공학박사 학위논문

Saliency Detection Methods Based on Multi-seed Propagation on Multilayer Graphs

다층 그래프에서의 씨드 전파를 통한

중요 객체 검출 방법

2016 년 8 월

서울대학교 대학원

전기컴퓨터공학부

황 인 성

ABSTRACT

Humans are able to understand visual information very quickly, even in complex scenes by unintentionally prioritizing the regions or objects in the scenes and spending their efforts to interpret things according to their priorities. Visual saliency detection is to find such important and noticeable things in images as humans do. It is exploited as a preprocessing tool of a variety of tasks in computer vision and image processing area. Thus, there have been many saliency detection methods which have achieved to locate salient objects in a certain degree. Recently, many researchers have focused on uniformly highlighting the salient objects via diverse approaches. In this respect, this dissertation presents novel methods for multilayer graph construction and multi-seed propagation to detect salient objects more accurately, which is applied to three topics: single image saliency detection, co-saliency detection of multiple images, and skin detection as a target specific task.

First, a single image saliency detection method is proposed based on the seed propagation approach on a graph. Unlike existing approaches, the proposed method exploits two different types of seeds for the propagation of saliency information, each of which stands for salient objects and background respectively. Two kinds of seeds are separately propagated to all nodes of a graph which is effectively learned by a semi-supervised learning scheme, and the saliency map is generated by combining

the results of each propagation. In addition, this approach is expanded to multi-layer graphs for better localization of salient objects homogeneously. Two different methods are presented in this dissertation, one of which maintains spatial coherence and another method focuses on increasing the consistency on feature space, e.g., color feature. Hence, both multilayer graph cases take advantage over the single layer graph case in a such way that these consider global and local relations of a whole image. Experiments demonstrate that introducing the multi seeds helps to reduce false positives and the performance is further improved by constructing the multilayer graphs. Moreover, the proposed approach outperforms the state-the-art methods in terms of various objective measures.

This dissertation also presents a method for co-saliency detection which aims at locating salient objects occurring in multiple images. Unlike saliency detection, it needs additional information to represent coherence of regions among images, so a pairwise coherence cue is designed to describe co-existence as well as saliency from similarity of saliency (SoS). The basic framework of the proposed method follows that of the proposed single saliency detection approach which mainly consists of multi seed extraction, graph construction and propagation steps. However, it is necessary that the graphs of each image are connected in order to take account of the saliency information of other images together and to propagate the seeds over the nodes of every image. The nodes between different images are indirectly connected via additional nodes (cluster nodes), because it is a very challenging problem to link those inter nodes directly when the scales and locations of salient objects are different among images in general. Experiments demonstrate that the co-saliency cues of each image are successfully transferred to all image nodes, and it is also shown that the proposed co-saliency detection method yields better results on objective measures

and visually plausible results compared to the state-of-the-art methods.

Lastly, the approach of multi-seed propagation on a multilayer graph is utilized in a specific classification task, i.e., skin detection which extracts skin pixels/regions in images, where the regions are mostly considered as salient when there exist humans in the images. Instead of saliency seeds, skin seeds are exploited along with background seeds for the propagation step, which can be achieved by adopting existing skin detectors. The graph can be connected more relevantly with skin information, because it is possible to use top-down features unlike most saliency detection methods. Experiments shows that the proposed method outperforms other methods including the classical skin detection method used for the seed extraction, which demonstrates that the proposed approach can be used for the skin detection, and it is expected that the proposed framework may also be utilized in other image segmentation/classification problems.

Key words: saliency detection, co-saliency detection, skin detection, multi-seed, seed propagation, multilayer graph

Student number: 2012-30239

Contents

List of Figures

List of Tables

Chapter 1

Introduction

All living things with eyes probably focus on the regions or objects at the first glance which are particularly different from other regions or background, or moving things. Specifically, when we watch TV or images, we prioritize the regions or objects and dedicate our efforts on them to understand the scene or objects. Saliency detection aims at finding such noticeable and important objects in the images as humans do, and it is an important task for improving the performance of image processing and computer vision tasks such as segmentation, classification [1], retrieval [2], retargeting [3, 4], image enhancement, content-based compression [5], video summarization [6] and so on.

In this dissertation, saliency detection approach based on multi-seed propagation on multilayer graphs is proposed. Unlike previous methods that exploit one type of seeds, the proposed approach employs two types of seeds for the propagation to overcome the problems of single type seed case. The proposed approach is expanded to co-saliency detection research by introducing pairwise coherence cue between regions. Further improvement for the proposed co-saliency detection is achieved by

connecting the graphs of each image together, which enables the seeds to be transferred beyond intra image. This approach is also applied to skin detection which is a target specific detection problem. It is verified from extensive experiments that the proposed approach overcomes the problem of a single seed case, and further increases detection accuracy of saliency, co-saliency and skin detection by introducing multilayer graphs.

1.1 Saliency Detection for Single Images

In this dissertation, a saliency detection method is proposed based on semi-supervised learning and multilayer graph representation of images, which is basically based on the random walk theory like many existing methods [7–13]. For example, saliency and background seeds for the graph representation and processing are obtained from

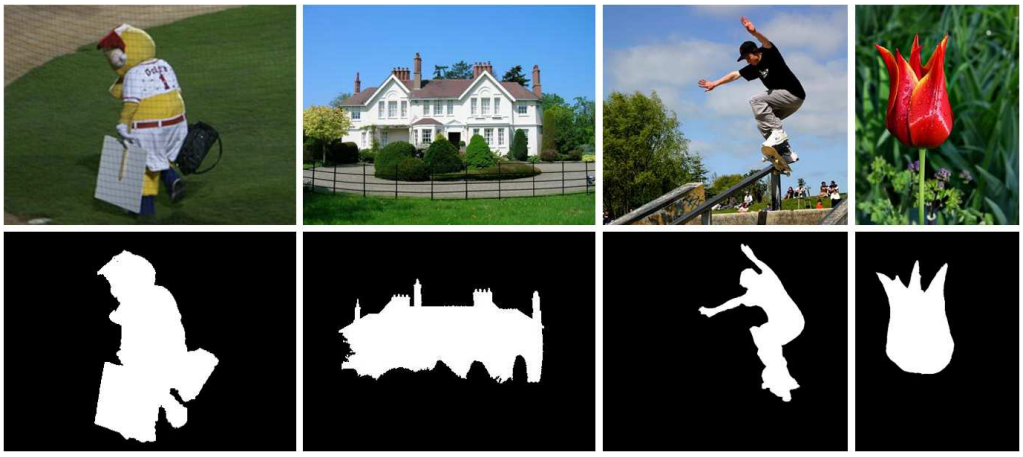


Figure 1.1: Example images containing saliency objects: top figures are example images, and bottom figures are binary mask images annotating salient objects, e.g, *Mascot*, *Mansion*, *Skateboarder* and *Flower*.

the hitting time in the view of random walkers in [7]. On the other hand, saliency seeds are extracted from the region-based global contrast, because the contrast is known to be one of the most powerful features for saliency detection [14, 15]. In addition, the center prior is combined with global contrast to enhance the performance of seed extraction. Also, the background seeds are extracted from the image boundaries, because the boundaries belong to mostly background regions as demonstrated in [16]. A semi-supervised learning is conducted to learn a graph for propagation referring to [12], which differs from the heuristic propagation methods [17] in that it considers local and global consistency. The two types of seeds are propagated to the learned graph from the observation that it is more effective than the case of using either of these seeds as in [10, 11]. Furthermore, the detection accuracy is improved by expanding the proposed approach to multilayer graph representation. Two multilayer graphs are introduced for regularizing the detection results, one of which is based on the idea that the multilayer graph can deal with various scales of objects as in [18, 19] and also deal with the spatial consistency, and another graph is to maintain the feature consistency, i.e., color features of regions. The multilayer graphs consequently make saliency values within the object more uniform and thus produce visually plausible results.

1.2 Co-saliency Detection for Multiple Images

Recently, co-saliency has emerged as an important subtopic of saliency detection, which is to find visually distinct regions and/or objects that commonly appear in a set of images. In other words, the co-saliency detection is to find commonly appearing saliency while suppressing salient objects/regions that appear in one or

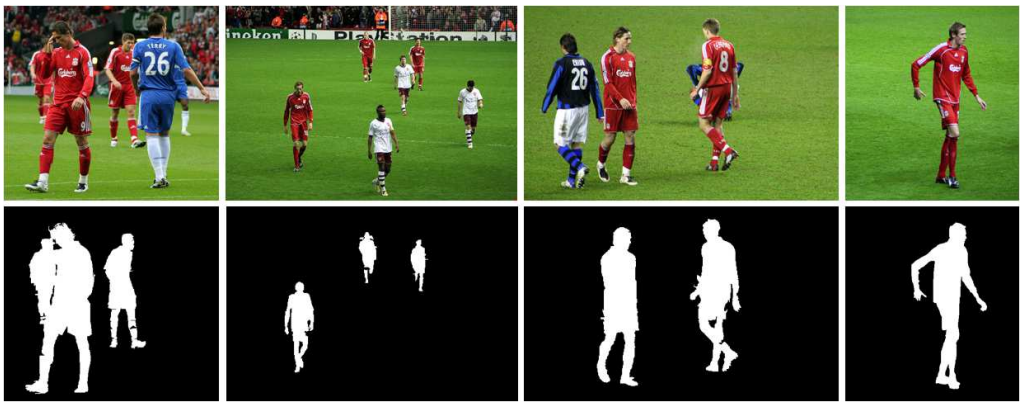


Figure 1.2: Example images containing co-saliency objects: top figures are a set of images including common saliency objects, and bottom figures are binary mask images annotating co-salient objects (*Red Soccer Players*).

a few images in the set. For the co-saliency detection, coherence among the images is additionally needed beside the cues used in saliency detection such as contrast prior [14,15,20] and/or boundary prior [11,16,20]. Co-saliency detection can also be applied to other tasks of computer vision, such as co-segmentation, co-recognition, image retrieval, image summarization and so on. It can be utilized in enhancing the single image saliency detection as well [21].

In this dissertation, a co-saliency detection method for multiple images is presented, based on a multilayer graph which multi seeds are propagated on. First, each image is over-segmented to form a set of regions and saliency values of segmented regions are computed from the contrast prior and center prior just like the proposed single image saliency detection approach. In addition, similarity of saliency (SoS) is defined to measure how each region is similar to salient objects. Co-saliency seeds are extracted based on both the saliency priors and SoS, while background seeds are decided to image boundaries. Image layers of each image are constructed by connect-

ing the over-segmented regions of each image separately, and cluster layer is built with cluster nodes generated from K-means clustering conducted with image nodes of all images. The image nodes of each image are independently connected to their intra nodes, and nodes of different images are indirectly connected via the cluster nodes. The multilayer graph is learned by the semi-supervised learning scheme [12], and the seeds are propagated to inter image as well as intra image, which takes advantage of utilizing seeds of different images. Co-saliency is computed from the propagation results of both seeds as in the saliency detection process.

1.3 Skin Detection

Skin detection is still regarded as a difficult problem due to illumination variations, skin color variations according to race and makeup, and the overlap between skin and non-skin pixels in the color space. This dissertation focuses on the third problem, i.e., the overlap of skin and non-skin regions in the color feature space by exploiting the spatial relationship of the regions. Consideration of spatial relationship in the image space is effective in discriminating the regions that are close in the feature space, but there are only a few methods that exploit this in skin detection problem [22–24].

In this dissertation, a new algorithm is presented, especially focusing on how to associate the spatial information with the skin detection problem. The proposed method is based on the graph representation of an image, where the graph is composed of two layers. One layer consists of image nodes which will be called image layer, where each node is a region of an over-segmented image. The other layer is called cluster layer, where the nodes are the clusters of image nodes according to their color similarity. The nodes in the image layer are connected according to their

spatial connectivity, and they are also indirectly linked to the cluster layer according to color similarity regardless of spatial distance. A learned graph is generated by the method of [12] and pre-defined seeds are propagated over the whole nodes of the graph, where two types of seeds are defined in the proposed method, i.e., skin and non-skin (background) seeds. Each kind of seed is separately propagated through the graph and a skin probability map is generated by combining the results of each propagation.

1.4 Contribution

Highlighting objects uniformly becomes very prominent in saliency detection recently. In this respect, this dissertation proposes a novel approach for localizing such objects with uniform degree from multi-seed propagation on well-designed multilayer graphs. The approach is utilized in saliency detection tasks, e.g., single image saliency, multiple image co-saliency and skin detection in the dissertation. Proposed methods detect salient objects more accurately in terms of various objective measures compared to other state-of-the-art methods. The contribution of the dissertation is summarized as:

- Introducing multi-seed for propagation, which reduces detection errors (especially false positive errors),
- Constructing multilayer graphs to consider global consistency of saliency maps,
- Providing a unified framework for saliency and co-saliency detection,
- Expanding to skin detection by substituting the procedure of seed selection and graph construction methodology.

1.5 Contents

In chapter 2, related works are presented which are the reviews of the conventional approaches for saliency, co-saliency, and skin detection. The proposed approach for the single image saliency detection is introduced in chapter 3. A saliency detection method with a single layer graph is explained in details, and it is expanded to two different multilayer based methods. Extensive experiments are followed by the explanation of the proposed approach. In chapter 4, a co-saliency detection method for multiple images is explained in details, and experimental results are also presented with discussion. The proposed approach is applied to a specific salient detection task: skin detection problem, and the detail explanations and experiment results are presented in chapter 5. Finally, the dissertation is concluded in chapter 6.

Chapter 2

Related Work

2.1 Saliency Detection

Many saliency detection methods have been proposed in several perspectives. Early works were based on biological and psychological observations where people are generally interested and gaze in the scenes, which is so called "fixation point" or "eye fixation" [25, 26]. Bruce *et al.* [27] proposed a saliency model based on Shannon's self-information to predict human fixation. Itti and Baldi [28] quantified surprise that measures how data affect an observer, and described a Bayesian definition of surprise for image saliency. Judd *et al.* [26] collected eye tracking data and learned saliency features to predict human attention. Meanwhile, many researchers used to define saliency as a visually noticeable or unique regions in images and focused on detecting such regions, rather than points. In other words, object level saliency detection has attracted more attention these days since it can be more attractive to be applied to other vision tasks.

Many saliency detection methods are based on bottom-up processing which is

data-driven and based on the prior knowledge of salient object and background. There are a wide variety of algorithms with so many priors for the saliency detection. For example, spectral residual and phase spectrum have been proposed to identify the salient regions in [29,30]. The center prior, which assumes that the salient objects are often located on the center of images, was also utilized in many literature [19]. Yang *et al.* [31] improved the center prior by introducing a convex-hull based center prior that estimates the location of salient objects as the interior of convex-hull. Also, contrast prior is regarded as one of the most powerful prior, because humans generally pay attention to high contrast regions. Itti *et al.* [32] proposed to detect saliency by measuring the local contrast as the center-surround difference of image patches in Gaussian pyramid images. Saliency detection algorithms based on global contrast have also been proposed in [14,33]. On the other hand, some literature have focused on background characteristics and a boundary prior was proposed from the observation that the salient objects rarely exist on image boundaries [8, 11, 16].

Recently, many algorithms represent the salient object to have uniform saliency value within the object (for finding salient regions rather than giving saliency points) while restraining all the unimportant backgrounds. Cheng *et al.* [34] proposed to use global color uniqueness and color spatial distribution, based on the image abstraction that divides an input image into several clusters from Gaussian mixture model (GMM) and component clustering. Also, the concept of group saliency detection has been proposed in [21], which retrieves similar images from large datasets and utilizes it together to refine the saliency. Gopalakrishnan *et al.* [7] exploited semi-supervised learning scheme using salient and non-salient nodes extracted from fully-connected graph and sparse graph. Yang *et al.* [11] similarly adopted a semi-supervised learning scheme, but their algorithm is based on ranking the nodes using either background

or salient nodes. Lu *et al.* [10] proposed an optimal seed extraction method for the semi-supervised learning. Yang *et al.* [31] applied initially detected saliency to graph-based regularization framework. Jiang *et al.* [8] introduced absorbing Markov chain in saliency detection. Kim *et al.* [18] proposed a multi-scale saliency detection method based on the random walk with restart. Yan *et al.* [19] unified saliency priors of each layer via an energy minimization framework.

2.2 Co-saliency Detection

Early works were focused on the detection of co-saliencies in a pair of images. For example, Jacobs *et al.* [35] captured local structure changes of co-salient objects in a little variation in the background. In [36], three different single-image saliency maps and multi-image saliency maps based on multilayer graph and SimRank optimization are integrated in a linear combination manner. In [37], superpixel affinity is computed based on graph matching between the images and propagated to entire images for co-saliency detection. The algorithm in [38] is based on rank-one constraint for fusing several single image saliency maps generated by different saliency detection methods.

Co-saliency detection algorithm for multiple images has been developed in a variety of aspects. Chang *et al.* [39] generated single image saliency maps for a set of images, and combined them with repeatedness score. In [40], the rank-one constraint method was expanded to the case of multiple images. Cluster-based method integrates contrast, spatial and corresponding cues in cluster level, and pixel level co-saliency is obtained from the similarity of pixels and clusters [41]. In [42], intra-image and inter-image saliency maps are generated based on multi-

scale segmentation and pairwise similarity ranking respectively, and these are mixed together for co-saliency detection. From hierarchically segmenting images, region contrast and similarity were defined in fine segmentation while object prior is defined in coarse segmentation, and these cues are combined to get co-saliency maps [43]. In [44], region and pixel similarity were exploited for region-level and pixel-level fusion. In [45], co-saliency was defined as the average probability of Gaussian mixture models (GMMs) which formulate foreground probability of each image. Two stage manifold ranking is conducted by exploiting the intra saliency maps of each image as quires [46].

In saliency and co-saliency detection, it is helpful to use segmentation/clustering as preprocessing steps, because it has advantage of exploiting some priors. In [43,44], the object prior applied in coarse segmentation played an important role in detecting co-saliency since coarse segmentation successfully separates objects and background. Cluster-based method [41] used the priors beyond intra images, and it makes co-salient regions in different image to be similar.

2.3 Skin Detection

There have been many researches for locating skin regions in an image for decades, and there are some literature on the appropriate color space for skin detection. Zarit *et. al* asserted that HS or HSV color space is the best for the skin detection based on the Bayesian theory [47]. Cheddad *et al.* proposed a new color space emphasizing the luminance channel for skin detection [48]. However, there are also some researches showing that the choice of different color space does not make a big difference in detection accuracy as long as the dimension of color space is same [49, 50].

In regard to the methodology for the discrimination of skin regions from the others, many kinds of skin description models were proposed. The simplest model may be to set a threshold or hyperplane that discriminates color features of skin in a certain color space [51, 52]. However, the accuracy of these detectors is limited because the decision boundary was determined heuristically. Some researchers proposed parametric skin models described with a few parameters, such as a single Gaussian model [53], a Gaussian mixture model [54] and an elliptical boundary models [55]. Although the parametric models can be trained with comparably small datasets, it is sometimes not enough to reflect the real skin color with these parameters. There are also some non-parametric skin detection methods that statistically models skin colors from plenty of images for training. For example, a Bayesian classifier was designed from histograms of training samples with annotation, and skin probability maps are generated from the Bayesian rule [56], and Khan *et al.* designed a skin detector based on the random forest approach [57]. The non-parametric methods have a little higher detection accuracy than the explicit and the parametric models, however, all of those methods are still sensitive to illumination changes, ethnic groups and training data.

Adaptive skin models have been proposed to overcome the above stated problems, which are shown to increase the detection accuracy especially by reducing false positive errors. Specifically, Zhu *et al.* proposed a two-stage skin detection method that extracts skin pixels by a classical method at first, and then adaptively learns a Gaussian mixture model [58] from the detected skin region in the first step. An illumination adaptation method was also proposed in [59], which selects one of skin models from the information of initial skin samples obtained from the Bayesian classifier and then fits the pre-trained model according to the samples. There are

also some methods that find skin region by using human-related features such as eye detector and face detectors, and then adjust the skin model based on the colors of pixels around these detectors [60, 61]. There is also a fusion method combining a dynamic threshold and a single Gaussian model [62], and a method based on a luminance adaptive color channel [63].

Recently, some researchers have focused on the spatial relation of skin pixels along with the skin colors that have been used before. For example, a controlled diffusion method was used to transfer the skin probability of pixels to its neighbors in [22]. A propagation of skin seeds based on the distance transform and a route optimization framework was proposed by [23], and they also modified their previous method by introducing a texture based seed extraction scheme [24]. Considering the spatial relations has contributed to overcoming the problem of overlapping skin and detecting non-skin pixels on the color spaces in a certain extent.

Chapter 3

Saliency Detection for Single Images

3.1 Proposed Approach for Saliency Detection

The proposed approach for saliency detection is presented in this section. First, an image is represented by a graph based on color affinity, where saliency seeds are extracted using a combination of color contrast and center prior, and background seeds are from boundary prior. Both seeds are propagated through the graph using a semi-supervised learning framework [12]. Then, the saliency of each node is computed from the propagated results. The overall procedure of the proposed method based on a single layer graph is illustrated in Figure 3.1. It is expanded to multilayer based methods, one of which is for spatial regularization and another is for color regularization respectively. Extensive experiments are presented to analyze the proposed approach, and the proposed methods are compared to the state-of-the-art methods to verify the performance.

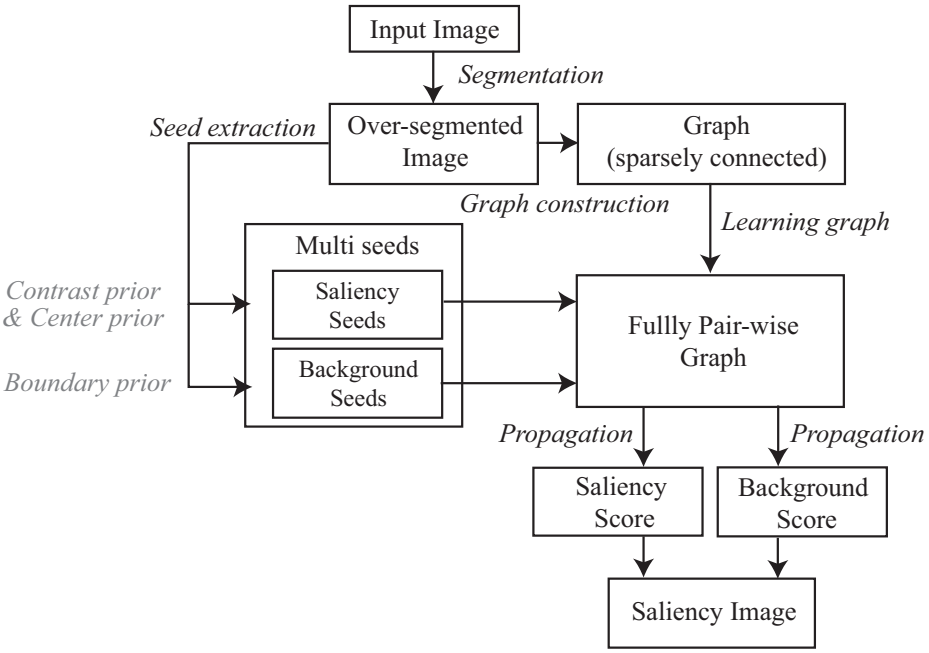


Figure 3.1: Flowchart of the proposed approach for saliency detection.

3.1.1 Graph Construction

First, the input image is divided into over-segment regions using the SLIC algorithm [64], which is a fast and memory efficient segmentation algorithm. Then a graph $G = (V, E)$ is constructed, where V is a set of nodes and E is a set of edges connecting the pairs of nodes. The nodes correspond to the over-segmented regions and the edge between the nodes v_i and v_j is given the weight w_{ij}, representing the affinity of two nodes. Color similarity is adopted as the affinity, which is written in an exponential form:

$$w_{ij} = \exp\left(-\left(\mathbf{x}_i - \mathbf{x}_j\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{x}_i - \mathbf{x}_j\right)\right), \tag{3.1}$$

$$\boldsymbol{\Sigma} = \frac{1}{|E|} \sum_{e_{ij} \in E} \left(\mathbf{x}_i - \mathbf{x}_j\right) \left(\mathbf{x}_i - \mathbf{x}_j\right)^T, \tag{3.2}$$

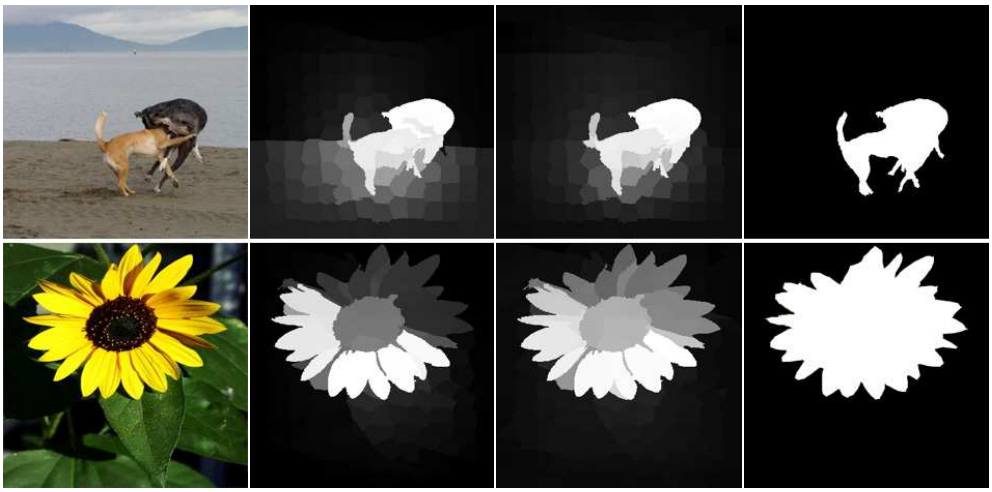


Figure 3.2: Comparison of results using the fixed parameter and the adaptive control parameter for the affinity (*from left to right*: input images, results of the fixed control parameter, results of the adaptive parameter and ground truth images).

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are average color vectors of nodes i and j respectively. Specifically, the pixels in the RGB space are transformed to the CIE Lab space, and $\mathbf{x}_n$ is defined as a vector *(L,a,b)* where each element is the average of corresponding elements of all the pixels in the node n. Also, $\mathbf{\Sigma}$ and $|E|$ are a color control matrix and the number of edges in the graph respectively. One of the contribution of the proposed algorithm is to introduce an adaptive parameter $\mathbf{\Sigma}$ to reflect image characteristics, as opposed to the fixed control parameter (σ) in the conventional methods [8,11,31]. In the case of using fixed control parameter, the background is not well suppressed in the case of low contrast image, while salient objects can be suppressed as well in the case of high contrast images as shown in Figure 3.2.

The nodes are connected only by their neighbors in the graph because locally connected graphs are more effective for graph learning to consider both of global and

local consistency simultaneously than the fully connected graphs which might lose local relationship of neighbors. In addition, local characteristics are finally propagated to the distant nodes so that many graph representations in image processing and vision algorithms consider local and global connectivity. However it is noted that connecting all the boundary nodes improves performance [8,11,16], because this structure considers geodesic distance which is especially meaningful for the boundary nodes. In summary, the graph for proposed algorithm is represented by an affinity matrix whose (i, j)-th element is defined as:

$$(\mathbf{W}_1)_{i,j} = \begin{cases} w_{ij}, & \text{if } j \in Q_i \text{ or } i, j \in B, \\ 0, & \text{otherwise,} \end{cases} \tag{3.3}$$

where Q_i is the neighbor of node i and B is the set of boundary nodes.

3.1.2 Seed Extraction

Saliency seeds

Two priors are exploited to extract saliency seeds. The first one is based on the contrast prior (RC [14, 15]), where a region saliency is defined as global contrast to all other nodes with spatial weight:

$$s_i^c = \sum_{j \neq i} \exp\left(-D_s(i,j)/\sigma_s^2\right) A(j) D_c(i,j), \tag{3.4}$$

where $D_s(i, j)$ and $D_c(i, j)$ are the spatial and color distance between the nodes i and j respectively, $A(j)$ is the area of the node j, and σ_s^2 is a spatial variance which is set to 0.4 as suggested in [15]. $D_c(i, j)$ represents the distance or contrast between the nodes i and j, and the exponential term and $A(j)$ in the above equation are the spatial weight and area weight respectively. The spatial weight is designed to

have higher value for the nearby nodes, and the area weight reflects how many pixels contribute to the calculation of contrast, which means that larger regions affect more than smaller ones in measuring the contrast as demonstrated in [14, 15] with many examples.

Another one is the center prior which assumes that salient objects are located at the center of images. It is represented by spatial distance between the node position and image center as:

$$s_i^s = \exp\left(-D_s(i, c)\right), \tag{3.5}$$

where c is the image center. Although the center prior sometimes gives negative effects on the detection accuracy when the salient objects are located close to the image boundaries, the overall detection accuracy is increased because most images do not have important objects around the boundary. Finally, the seed saliency values are obtained by multiplying the contrast prior and the center prior,

$$s_i = s_i^c \times s_i^s. \tag{3.6}$$

The contrast and center prior can be easily calculated with a small amount of computation because the regions are already segmented when the graph is constructed. Examples of saliency map using the priors will be shown in Figure 3.5 in section 3.3, where the effectiveness of priors is also demonstrated. The contrast prior plays a major role for extracting saliency seeds while the center prior helps the background suppression.

Top $(100\times\gamma)\%$ of salient regions are extracted as saliency seeds. They sometimes contain high contrast clutter as well as salient nodes, which may affect detection accuracy when the clutter regions are used in the propagation step. Nevertheless, the final result is not much affected by the clutter regions, unless there exist a

significant amount of clutter in the saliency seeds. It will be discussed in section 3.3.

Background seeds

Since the nodes on the image boundary have little chance of being a part of salient objects, boundary nodes are adequate to background seeds. When the background seeds are from high contrast regions which may belong to salient nodes, it is contradictory and thus the seeds are excluded from both salient and background seeds. In summary, saliency and background seeds are defined as:

- Saliency seeds ($\mathbf{y}_{s1}$) : high contrast nodes which are not on the image boundary.

- Background seeds ($\mathbf{y}_{b1}$) : low contrast nodes on the image boundary.

3.1.3 Seed Propagation

As stated previously, both of saliency and background seeds are utilized for the semi-supervised learning. A graph-based learning method is adopted which is proposed by Zhou *et al.* [12], which can also be interpreted as a learning full pairwise graph problem [9]. The learned full pairwise graph of the first layer graph $\mathbf{W}_1$ in equation (3.3) is written as:

$$\mathbf{W}_{L1} = (1-\alpha)\left(\mathbf{D}_1 - \alpha\mathbf{W}_1\right)^{-1}, \tag{3.7}$$

where α is a learning balance parameter and $\mathbf{D}_1 = diag\left\{d_1, d_2, \cdots, d_N\right\}$ is the degree matrix of $\mathbf{W}_1$, whose diagonal entry is the degree of each node ($d_i = \sum_j w_{ij}$). Then the k-th column of $\mathbf{W}_{L1}$, denoted as $\mathbf{w}_{L1}^k$, is the learned affinity of the node k to

other ones, which is also the solution of minimizing the following energy function:

$$\mathbf{w}_{L1}^{k} = \arg\min_{\mathbf{z}} \sum_{i=1} d_i \left| z_i - \frac{c_{ik}}{d_i} \right|^2 + \lambda \sum_{i,j=1} w_{ij} \left| z_i - z_j \right|^2, \tag{3.8}$$

where c_{ik} is 1 for $i = k$, 0 for $i \neq k$, $\mathbf{z} = [z_1, \cdots, z_N]^T$ is an affinity vector for node k, and λ is a regularization parameter. The first term of equation (3.8) is the fitting term and the second is a regularization term which enforces two nodes with high affinity to have similar saliency values.

To apply this propagation method to the proposed saliency model, overall affinity after propagation of the saliency and background seeds is calculated which is written as:

$$\mathbf{f}_{s1} = \mathbf{W}_{L1}\ \mathbf{y}_{s1} = \sum_{i \in S_s} \mathbf{w}_{L1}^{i}, \tag{3.9}$$

$$\mathbf{f}_{b1} = \mathbf{W}_{L1}\ \mathbf{y}_{b1} = \sum_{i \in S_b} \mathbf{w}_{L1}^{i}, \tag{3.10}$$

where S_s and S_b represent saliency and background seed sets respectively. Finally, the saliency is defined as:

$$\mathbf{s}_1 = (\mathbf{f}_s - \mathbf{f}_b)\,./\,(\mathbf{f}_s + \mathbf{f}_b), \tag{3.11}$$

where ./ is the element-wise division of two vectors. The numerator represents the saliency while the denominator adjusts the balance of saliency among the nodes. Finally, this is normalized to $[0,\ 1]$.

3.2 Expanding to Multilayer Graph Based Saliency Detection

A multilayer graph is exploited to further increase the performance of the proposed saliency model. The introduction of the multilayer graph helps saliency maps more

consistent in terms of spatial space or feature space, and the saliency maps are more adequate to be utilized in various applications. Two different types of regularization frameworks are introduced in this dissertation. First method is a spatial regularization framework which concerns interlayer connection based on spatial correlation, and another method is a color regularization framework which considers color relations between inter nodes. Both methods are explained in following subsections.

3.2.1 Method 1: Spatial Regularization Framework

The image is multiply segmented by SLIC [64] algorithm by varying segment size, and a multilayer graph connecting both intra and inter nodes are constructed as shown in Figure 3.3. Let $\mathcal{I}_m$ be the m-th layer segmented image, where small m means finely segmented layer. Intralayer nodes are then connected with the same method as the single layer case defined in equation (3.1). In the case of interlayer connection, only spatial relation is considered to liking edges. Edge strength for interlayer is designed to be proportional to its intersection area, i.e., the connection weight is defined as,

$$w_{ij}^{pq} = \epsilon \cdot \frac{A\left(r_i^p \cap r_j^q\right)}{A\left(r_i^q\right)}, \quad q > p \tag{3.12}$$

where p and q indicate layer indexes, r_i^p is the region of node i in the p-th layer, $A(\cdot)$ is the area of region, and ϵ is an inter-weight control parameter. It means that there is no connection between the inter nodes if they do not share image pixels. The multilayer affinity matrix can then be represented as a block matrix. For example, in a two layer graph case, it can be written as:

$$\mathbf{W}_2 = \begin{bmatrix} \mathbf{W}^{11} & \mathbf{W}^{12} \\ \mathbf{W}^{21} & \mathbf{W}^{22} \end{bmatrix}, \tag{3.13}$$

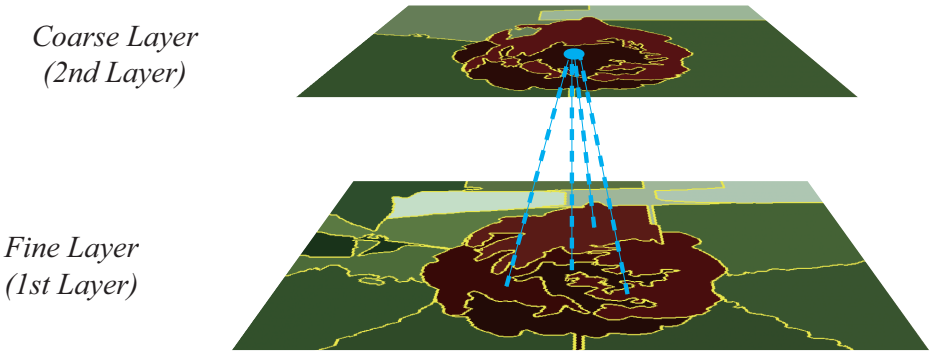


Figure 3.3: An illustration the inter layer connection for spatial regularization (*blue dot* lines: inter layer edges between a coarse layer node and its corresponding nodes of the fine layer).

where $\mathbf{W}^{11}$, $\mathbf{W}^{22}$ are intralayer affinity matrices and $\mathbf{W}^{12}$, $\mathbf{W}^{21}$ are interlayer affinity matrix. Note that $\mathbf{W}^{11} = \mathbf{W}^{1}$ in equation (3.3), and $\mathbf{W}^{21}$ is the transpose of $\mathbf{W}^{12}$.

Now, saliency detection is formulated in the m-layer graph representation. The saliency and background seeds are only extracted from the first layer, i.e.,

$$\mathbf{y}_s = \left[\mathbf{y}_{s1}; \mathbf{0}; \cdots; \mathbf{0}\right], \quad \mathbf{y}_b = \left[\mathbf{y}_{b1}; \mathbf{0}; \cdots; \mathbf{0}\right], \tag{3.14}$$

because nodes in the higher layer correspond to coarsely segmented regions which may contain both background pixels and object pixels simultaneously.

The seeds are propagate to all the nodes similar to equation (3.9) and (3.10) except that $\mathbf{W}_m$ and $\mathbf{D}_m$ are used instead of $\mathbf{W}_1$ and $\mathbf{D}_1$ to obtain the overall affinity to the seeds $\mathbf{f}_s = [\mathbf{f}_{s1}; \cdots; \mathbf{f}_{sm}]$ and $\mathbf{f}_b = [\mathbf{f}_{b1}; \cdots; \mathbf{f}_{bm}]$ where m is the number of layers. Since the seeds are distributed to all the layers after the propagation, it is needed to be gathered to compute the final saliency. The distributed seeds are collected at the first layer which is the finest layer. It can be done by moving the seeds of higher layers to the first layer according to the transition matrix between

the layers, and adding them together. It is written as:

$$\begin{aligned}\hat{\mathbf{f}}_{s1} &= \mathbf{f}_{s1} + \sum_{i=2} \left(\mathbf{P}^{i1}\right)^T \mathbf{f}_{si} \\ &= \mathbf{f}_{s1} + \sum_{i=2} \mathbf{P}^{1i} \mathbf{f}_{si},\end{aligned} \tag{3.15}$$

$$\begin{aligned}\hat{\mathbf{f}}_{b1} &= \mathbf{f}_{b1} + \sum_{i=2} \left(\mathbf{P}^{i1}\right)^T \mathbf{f}_{bi} \\ &= \mathbf{f}_{b1} + \sum_{i=2} \mathbf{P}^{1i} \mathbf{f}_{bi},\end{aligned} \tag{3.16}$$

where, $\mathbf{P}^{ij}$ is the transition matrix between i-th layer and j-th layer, which is the column normalized matrix of $\mathbf{W}^{ij}$. Final saliency is computed using equation (3.11).

3.2.2 Method 2: Color Regularization Framework

The segments of the fine layer correspond to image layer nodes and a cluster layer is introduced for color regularization. It is noted that only one image layer is necessary for the color regularization framework as well as the cluster layer. All color features in the segments are grouped to make clusters using K-means clustering algorithm, and each of the cluster is set to the nodes of cluster layer. The feature for the cluster node i is notated as $\mathbf{c}_i$ which is the average of color vectors of image nodes belonging to the cluster i and C_i is a set of the corresponding image nodes. The cluster nodes are connected not only with themselves for intra graph connection but also with the image layer nodes for interlayer graph connection as shown in Figure 3.4. In the intralayer connection for the clusters, each cluster node of the graph is sparsely connected to its k nearest neighbor (k-NN) in the color feature space. Edge strength for the intralayer is designed with color similarity between nodes, which is written as:

$$w_{ij}^{CC} = \exp\left(-\frac{\|\mathbf{c}_i - \mathbf{c}_j\|_2}{\sigma_c}\right), \tag{3.17}$$

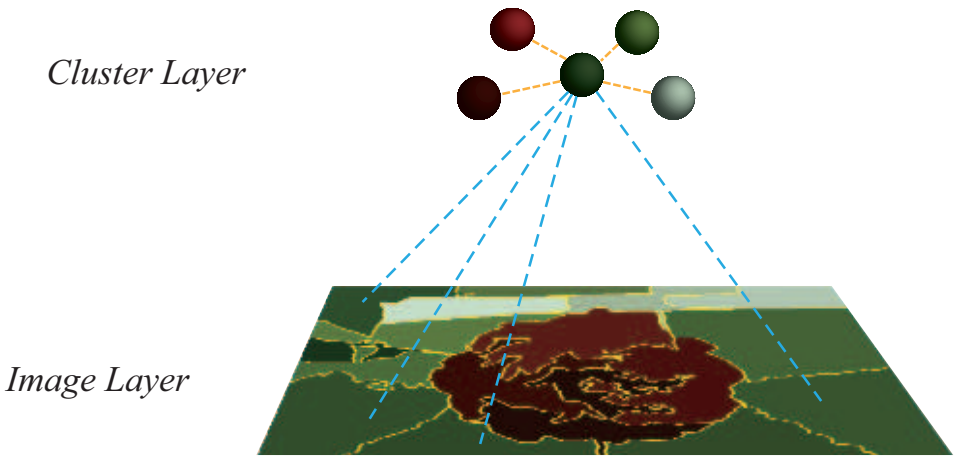


Figure 3.4: An illustration of the inter layer and intra cluster layer connections for color regularization (*blue dot lines*: inter layer edges between a cluster layer node and its corresponding nodes of the image layer, *yellow dot lines*: intra cluster layer edges).

$$(\mathbf{W}^{CC})_{i,j} = \begin{cases} w_{ij}^{CC}, & \text{if } i \in k\text{-NN}(j) \text{ or } j \in k\text{-NN}(i), \\ 0, & \text{otherwise}, \end{cases} \tag{3.18}$$

where $\mathbf{c}_i$, $\mathbf{c}_j$ are the color features of cluster node i and j, σ_c is a color control parameter for the intra cluster connection, and k-NN(i), k-NN(j) are sets of k-nearest neighbors of the cluster node i and j respectively.

Meanwhile, in the case of interlayer connection which represents the relation between image nodes and cluster nodes, each image node is connected to only one cluster node which has the smallest distance to the image node, and it is determined in the process of K-means clustering. Similar to equation (3.17), weight of edges between image nodes and cluster nodes are defined based on color similarity as below:

$$w_{ij}^{1C} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{c}_j\|_2}{\sigma_c}\right), \tag{3.19}$$

$$(\mathbf{W}^{1C})_{i,j} = \begin{cases} w_{ij}^{1C}, & \text{if } i \in \mathbf{C}(j) \\ 0, & \text{otherwise,} \end{cases} \tag{3.20}$$

where $\mathbf{x}_i$, $\mathbf{c}_j$ are the color feature of image node i and an cluster node j respectively, and $\mathbf{C}(j)$ is a set of image nodes belonging to cluster j.

The multilayer affinity matrix is expressed as a block-wise matrix, similar to equation (3.13) and it can be written as:

$$\mathbf{W}_3 = \begin{bmatrix} \mathbf{W}^{11} & \mathbf{W}^{1C} \\ \mathbf{W}^{C1} & \mathbf{W}^{CC} \end{bmatrix}, \tag{3.21}$$

where $\mathbf{W}^{11}$, $\mathbf{W}^{CC}$ are intralayer affinity matrices for image nodes and cluster nodes respectively, and $\mathbf{W}^{1C}$, $\mathbf{W}^{C1}$ are interlayer affinity matrices where $\mathbf{W}^{C1}$ is the transpose of $\mathbf{W}^{1C}$.

All the procedure of saliency detection are identical to those of the method 1 stated in section 3.2.1 except that the color regularization framework uses $\mathbf{W}_3$ instated of $\mathbf{W}_2$, thus it is omitted in this subsection.

3.3 Experiments

In this section, extensive experimental results are presented which have been conducted on famous datasets: ASD [33, 65], MSRA10K [15, 65], DUT-OMRON [11] and ECSSD [19]. The parameters used in the proposed methods are determined which the proposed methods are conducted with in all experiments. For evaluation, various objective measures are used to evaluate the performance of the algorithms, and some results images are provided for intuitive visual comparison. Firstly, experiments are carried out on ASD dataset in order to analyze the internal structure

of the proposed methods, e.g., effectiveness of priors, seeds and layers. Then, the proposed methods are compared with existing methods on entire datasets.

3.3.1 Evaluation Measures

Four widely used criteria are adopted to evaluate quantitative performance: precision-recall curve (PR curve), ROC curve, area under PR and ROC curve (AUC-PR, AUC-ROC) [66, 67]. Additionally, precision, recall and F-measure (PRF) in a fixed threshold are evaluated. Precision and recall are referred as positive predictive value (PPV) and true positive rate (TPR) respectively, and these are defined as:

$$Precision = \frac{TP}{TP + FP}, \tag{3.22}$$

$$Recall = \frac{TP}{TP + FN}, \tag{3.23}$$

where TP, FP and FN are the number of true positive, false positive and false negative samples. Saliency maps are binarized by a threshold which is varying from 0 to 255, and precision and recall at each threshold are computed by comparing the map with the ground truth images. PR curve is plotted from precision and recall values of each threshold. Similarly, ROC curve is plotted by calculating false positive rate and true positive rate, where false positive rate is defined as:

$$FPR = \frac{FP}{FP + TN}. \tag{3.24}$$

As overall measures of the curves, AUC is additionally computed by accumulating areas under PR and ROC curve. A maximum F-measure is evaluated and corresponding precision and recall values are also demonstrated in PRF to compare performance of binarized detection results, where general F-measure is defined with

precision and recall:

$$F_\beta = \frac{\left(1+\beta^2\right) Precision \times Recall}{\beta^2 Precision + Recall}. \tag{3.25}$$

where β^2 is set to 0.3 in order to emphasize the importance of precision as suggested in [33].

3.3.2 Experiment Setup

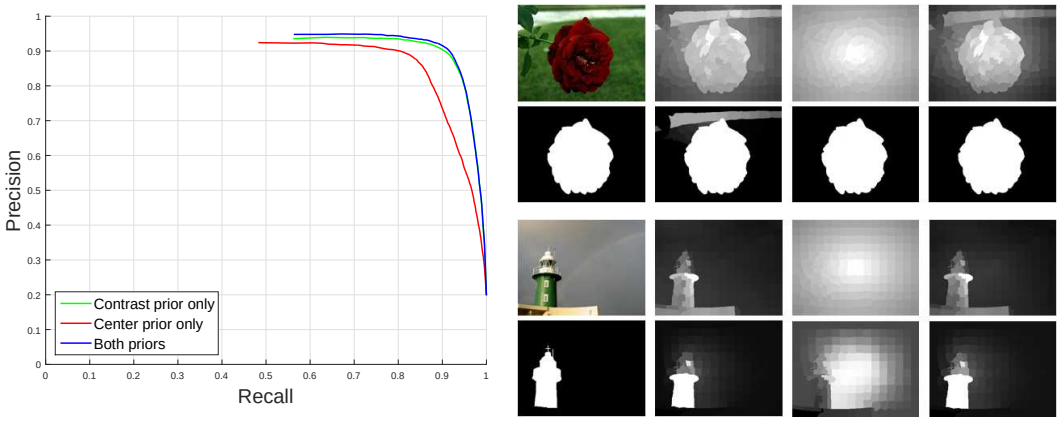
The number of layer m and inter-weight control parameter ϵ are set to $m = 2$ and $\epsilon = 1.0$, empirically, the selection of which will be discussed in section 3.3.3. The number of super-pixels (nodes) in the first layer is set to $N_1 = 250$ and the number of super-pixels in the second layer is set to $N_2 = 50$, which is 20 percent of N_1 that is appropriate for connecting neighbor super-pixels. The learning balance parameter α is set to 0.99 as suggested in [12], where larger α yields more regularized results. About the γ, which is the ratio of extracting salient nodes out of the overall nodes, it is set to be less than the ratio of saliency pixels over the overall pixels, because false positive seeds negatively affect more than false negative seeds. Since the ratio of saliency pixels is between 0.15 and 0.30 for the four datasets experimented in this chapter, γ is set to 0.1 for preventing false positive seeds.

3.3.3 Analysis of the Proposed Approach

The proposed methods can be varied depending on the structures of the algorithms, which is investigated in this subsection. Analysis is provided with respect to effectiveness of priors, seeds and layers. For this, results of various situation are presented in terms of PR curves on ASD dataset. the visual comparisons are also provided for explanation.

Effectiveness of saliency priors

As explained in section 3.1.2, two priors (contrast and center prior) are adopted to extract the saliency seeds. PR curves and saliency detection results obtained with each of the priors are shown in Figure 3.5a, where it can be seen that the results using the contrast prior is much better than that of the center prior only case, because the contrast is a very powerful feature to detect salient regions [14, 15]. Hence, the contrast prior plays a major role in the proposed algorithm while the center prior helps to increase the performance slightly. Because people tend to take pictures with the salient object near the image center, the overall performance is increased by multiplying the center prior and the contrast prior. The role of center prior is



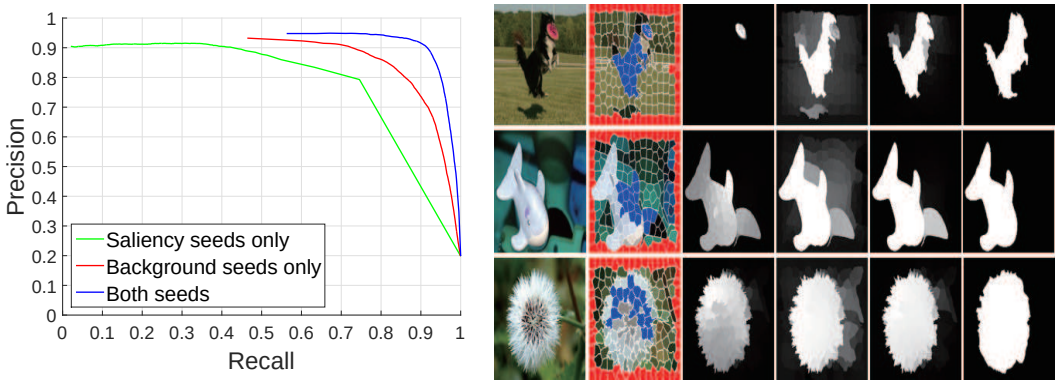
(a) (b)

Figure 3.5: Comparison of the priors on ASD dataset: (a) PR curves, (b) visual comparison (*first row*: input images, results of the contrast prior, results of the center prior, and results of both priors, *second row*: ground truth images, propagation results with the contrast prior only, propagation results with the center prior only, and propagation results with both priors).

to weaken the high contrast region in the boundaries, which contributes to rejecting background regions with high contrast as the *Flow* image demonstrated in Figure 3.5b. Meanwhile, the center prior does not negatively affect the detection results in the case that salient objects are not in the center as the *Lighthouse* image in Figure 3.5b, if no high contrast background is around the image center.

Effectiveness of Seeds

In this experiment, the effect of choosing the seeds for the propagation stage is examined, i.e., effect of (a) using only saliency seeds, (b) using only background seeds, (c) using both seeds. In the case of using only saliency seeds, salient objects are not uniformly highlighted or even missed as shown in Figure 3.6b. Meanwhile, it fails to suppress background regions when background seeds are solely used. When both kinds of seeds are exploited, it can detect the salient objects uniformly and suppress background regions as well. It means that proposed method successfully takes the advantages of multi-seeds, i.e., saliency seeds and background seeds. For the quantitative comparison, Figure 3.6a shows the PR curves which supports that using both kinds of seeds outperforms the other cases.



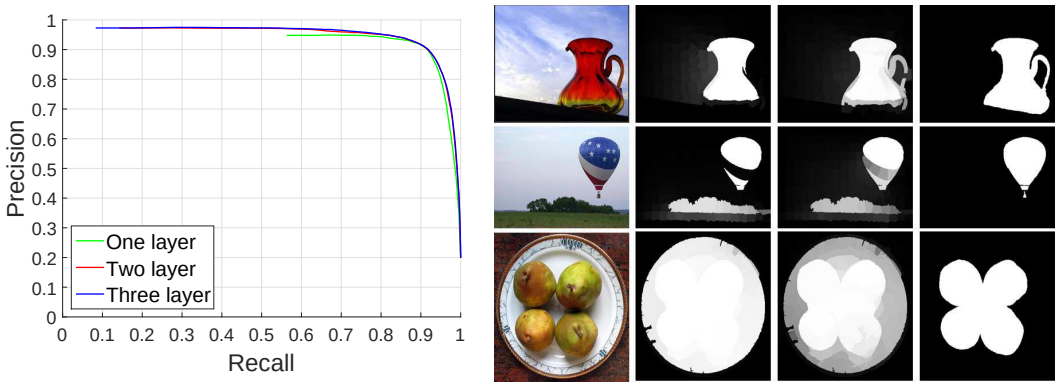
(a) (b)

Figure 3.6: Comparison of saliency and background seeds on ASD dataset: (a) PR curves, (b) visual comparison (*from left to right*: input images, extracted seeds images, results with saliency seeds only, results with background seeds only, results with both seeds, and ground truth images).

Effectiveness of Layers

The effect of layer is examined in this subsection. The motivation of using multilayer graph is to alleviate the problems of single layer graph that are found through the experiments. The first problem is that it fails to detect some parts of objects or interior of objects when these parts have a bit different color from that of majority salient parts (1-2 rows in Figure 3.7b). The second problem is that it tends to detect regions with the extreme saliency values around 0 or 1 (too much binarized) as shown in 1st and 3th row of Figure 3.7b. It sometimes gives background with high saliency values, which causes false positive detection in low recall ranges and cannot satisfy certain degree of recall value as demonstrated in Figure 3.7a.

For the spatial regularization framework (Method 1), the higher layer works with larger areas and thus complements the weakness of the single layer graph on



(a) (b)

Figure 3.7: Comparison of the number of layers on ASD dataset: (a) PR curves, (b) visual comparison (*from left to right*: input images, results of one layer case, results of two layer case, and ground truth images).

Table 3.1: Comparison of Precision, recall, F-measure and AUC with different ϵ on ASD dataset.

ϵ	**Precision**	**Recall**	**F-measure**
0.01	0.9249	0.8726	0.9123
0.1	0.9317	0.8648	0.9153
0.5	0.9346	0.8656	0.9177
1.0	0.9388	0.8524	0.9182

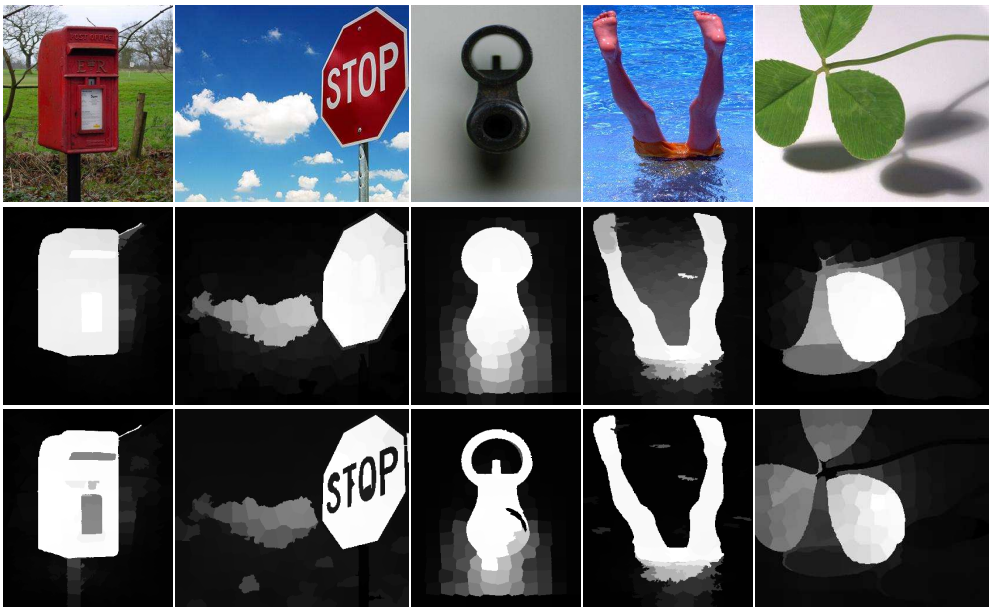


Figure 3.8: Comparison of two regularization frameworks (*from top to bottom*: input images, results of the spatial regularization framework and results of the color regularization framework).

the over-segmented image. According to extensive experiments, the second layer improves the performance but the third one ($N_3 = 25$) does not any more as shown in Figure 3.7a, and hence only two layer graph are used in all the experiments. In addition, the effect of the inter-weight control parameter ϵ is examined by varying it from 0 to 1. Table 3.1 shows that it performs better when ϵ is larger, because larger ϵ yields more regularized results in general.

The color regularization framework, Method 2, considers feature correlation in the regularization framework which enforces that the nodes sharing similar features keep similar saliency, while Method 1 focuses on making saliency maps spatially uniform. There are pros and cons in both methods. Method 1 have an advantage over Method 2 in that it detects saliency more likely to objects since the parts of

objects have similar saliency values. However, the method 1 can suppress the small salient regions far from large salient objects because background regions beside those regions contribute to reducing the saliency values, while method 2 is capable of keeping those regions with higher saliency values as demonstrated in Figure 3.8. In summary, method 1 is better to capture salient objects while method 2 is superior in detecting saliency in the perspective of feature space.

3.3.4 Comparison with Other Algorithms

Experiments are carried out on ASD [33,65], MSRA10K [15,65], DUT-OMRON [11] and ECSSD [19]. ASD and MSRA10K are relatively simple whereas DUT-OMRON and ECSSD are complex scene datasets. The results of those two groups are individually presented along with discussion. The proposed methods are compared with some of the most cited methods (IT [32], SR [30], FT [33], AC [68], GB [69]) and the best performing methods based on contrast or boundary prior (RC [15], GR [31], HS [19], GMR [11] and MC [8]). It is noted that Method 1 and 2 are notated as the proposed MM1 and MM2 respectively.

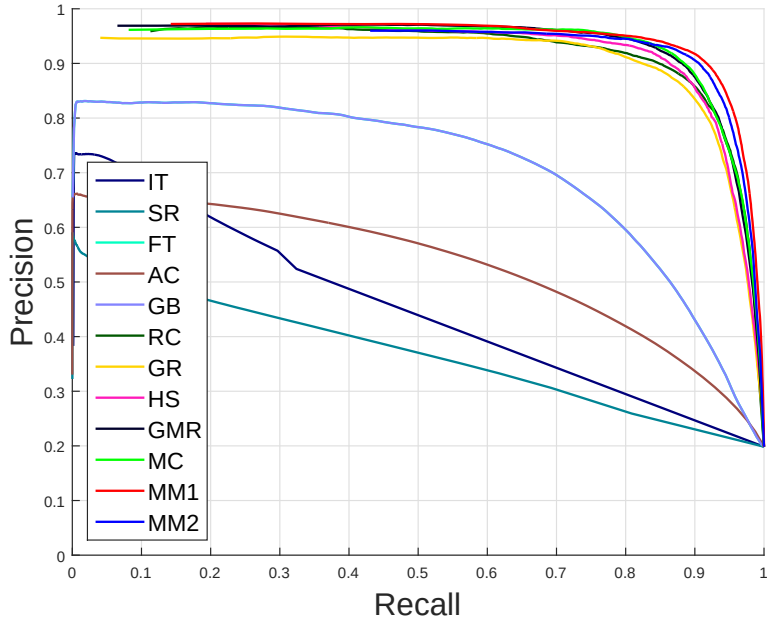
The IT computes saliency from the local contrast defined as the center-surround difference of image patches in Gaussian pyramid images. The SR focuses the saliency in spectral domain. The FT defines the saliency as color difference between each of the pixels and average value on the entire image, which can be interpreted as global color contrast. The GB introduces the graph representation in the saliency detection. RC, GR and HS exploit the contrast prior while GMR and MC are based on the boundary prior. The RC is a global color contrast algorithm based on segmented regions and it is recently modified and improved in [15]. The GR is based on convex-hull prior as well as global contrast. A graph based regularization framework is

applied for uniformly detecting salient objects and suppressing background regions in GR. In HS, the multilayer saliency unifies saliency cues of each layer using an energy minimization framework, where each layer's saliency is based on the contrast and center prior. The GMR ranks the similarity of regions with background cues extracted from image boundaries. MC measures the time that every random walker moves to absorbing nodes which are duplicated from image boundaries.

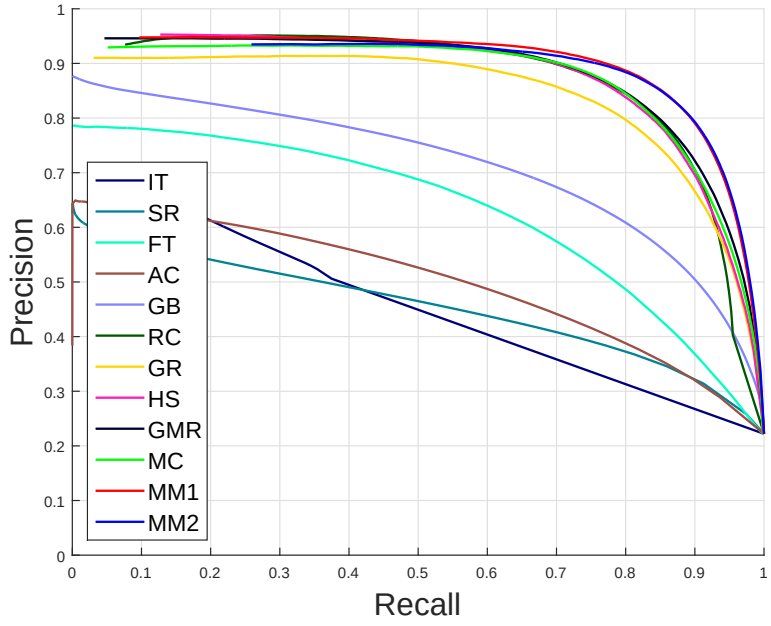
Simple sets

Experiments are conducted on simple datasets in this subsection. ASD, a subset of MSRA dataset [65], consists of 1,000 natural images with unambiguous salient objects. It is provided with accurate annotation images by [33], and the most popular dataset for saliency evaluation. MSRA10K is the largest dataset in saliency detection research which contains 10,000 images along with annotation images which is also randomly selected from MSRA dataset by Cheng *et al.* [15]. Each image of both sets contains a single salient object which is located in the center of the image, so it is regarded as relatively simple dataset.

It can be seen from PR curevs and ROC curves (Figure 3.9 and Figure 3.10) that the methods focusing on regularization of saliency map, i.e., GB, RC, GR, HS, GMR MC and the proposed methods, are a far superior to classical methods depending mostly on saliency priors on both ASD and MSRA10K datasets. So, those elaborate methods can localize salient objects much more efficiently than the classical methods. These figures also show that the two proposed methods outperform other methods where the curves of the proposed methods are placed higher in overall ranges. Especially, the proposed method based on spatial regularization framework works slightly better than the color regularization method. The reason is that

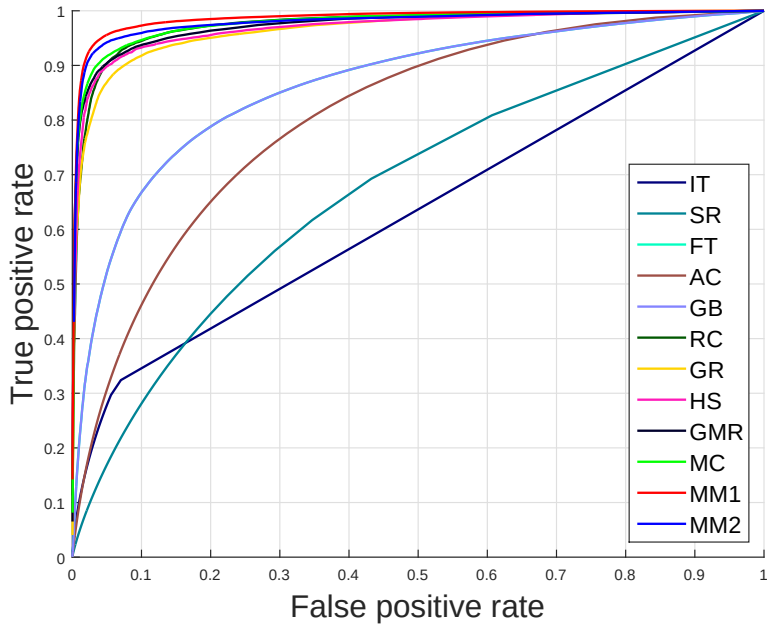


(a)

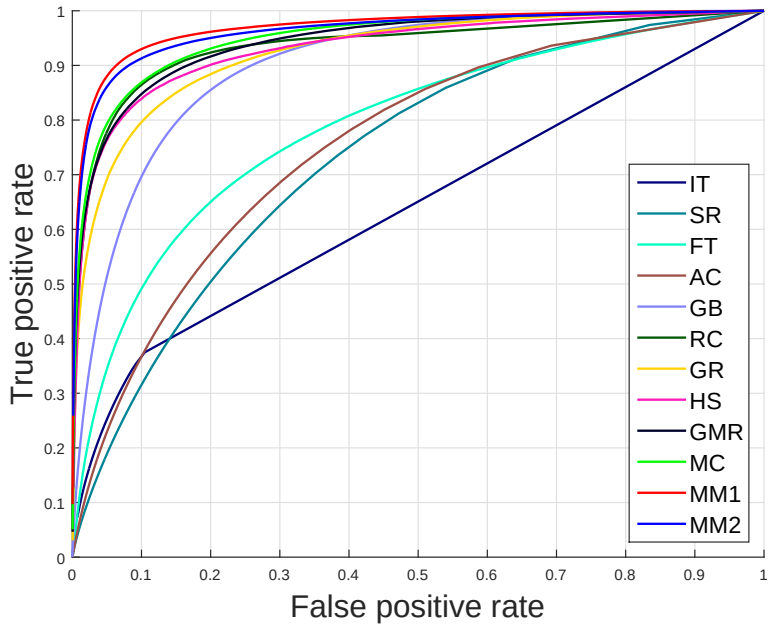


(b)

Figure 3.9: Comparison of PR curves on (a) ASD and (b) MSRA10K datasets.



(a)



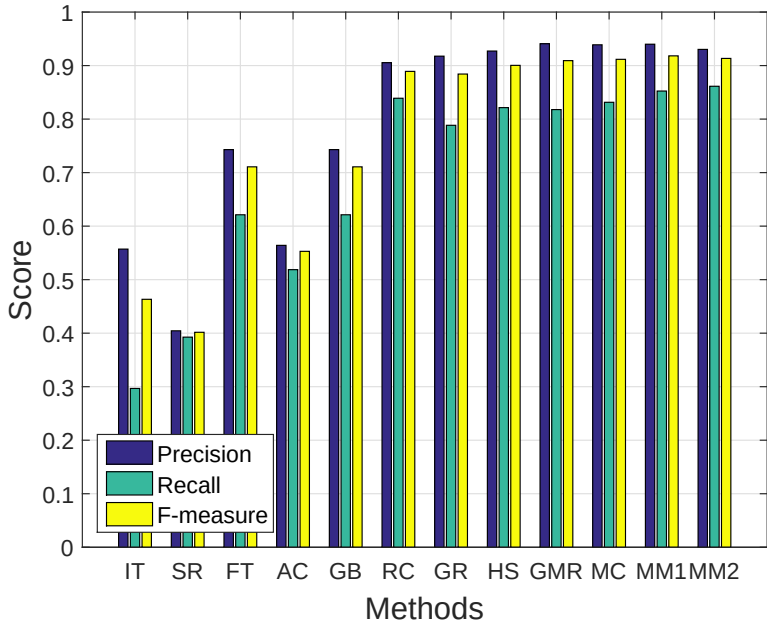
(b)

Figure 3.10: Comparison of ROC curves on (a) ASD and (b) MSRA10K datasets.

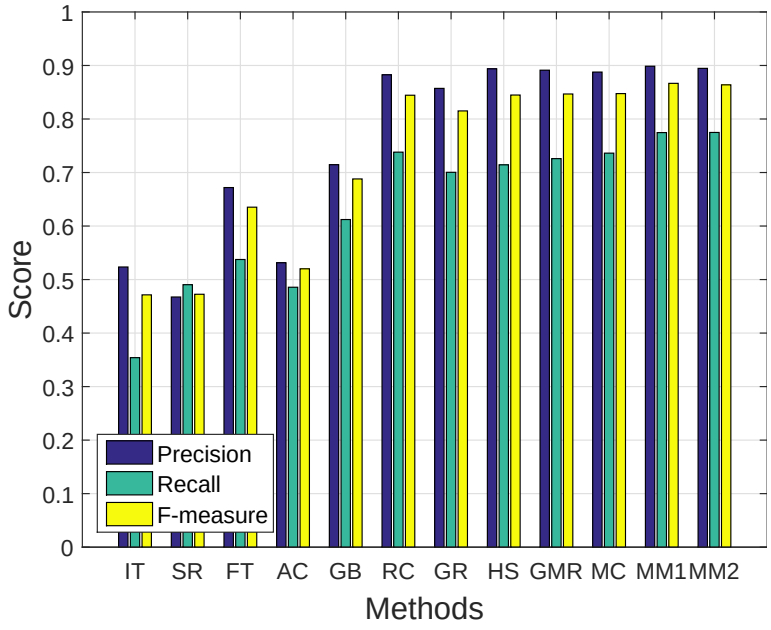
the color regularization method is designed to suppress color regions which differ from salient objects and highlight similar colors regions in a certain degree even if the regions do not belong to salient objects as shown in 5th and 6th columns of Figure 3.12. Meanwhile, the spatial regularization method is to locate salient objects uniformly considering the spatial relation, so it has an advantage in the single object case as ASD and MSRA10K datasets. Overall scores for both curves are presented in terms of AUCs in Table 3.2 and Table 3.3 which also demonstrate that the proposed methods produce better results in general.

An additional experiment is conducted to evaluate the performance of binarized saliency results with a certain threshold which is determined to maximize F-measure of each algorithm. It can be seen from Figure 3.11 that both the proposed methods record first and second places in terms of F-measure on ASD and MSRA10K datasets though the differences between the proposed methods and the other methods based on regularization are not significant. It implies that the proposed methods have also a little merit of being applied in other tasks in the form of the binary map.

Visual comparison is illustrated in Figure 3.12 which shows that the proposed spatial regularization method generally highlights salient object in a uniform degree while other methods often fail to detect the salient object consistently. Though the proposed color regularization method comparably localizes salient objects uniformly than existing method, it remains some regions undetected whose color is quite different from major color of salient objects e.g., *Flag* and *Beer* images. In the view of background suppression, though the proposed methods sometimes fail to suppress the background regions (*Beer* image) due to excessive regularization, the proposed methods show comparable or better results compared to others.



(a)



(b)

Figure 3.11: Comparison of precision, recall and F-measure at peak F-measure on (a) ASD and (b) MSRA10K datasets.

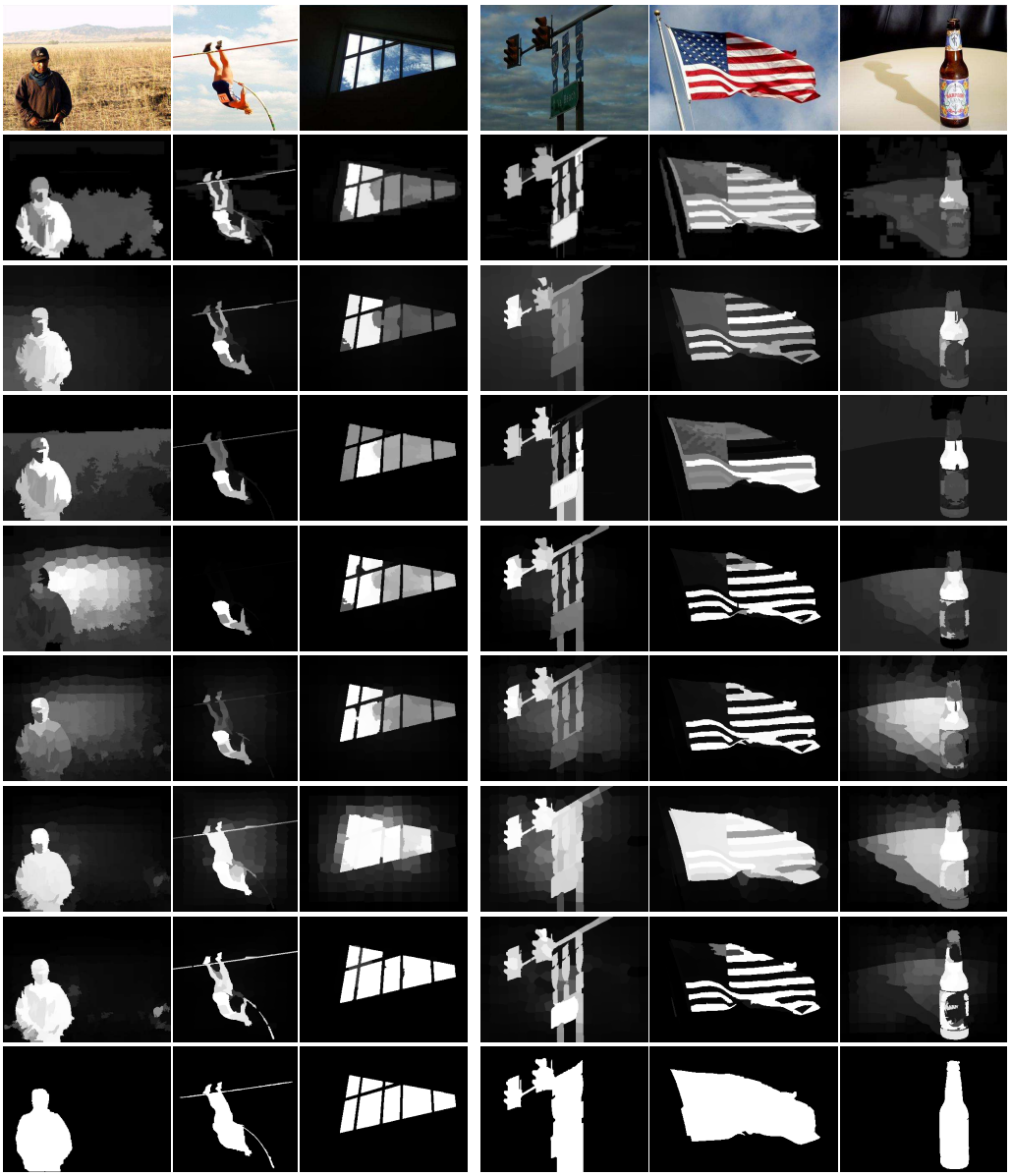
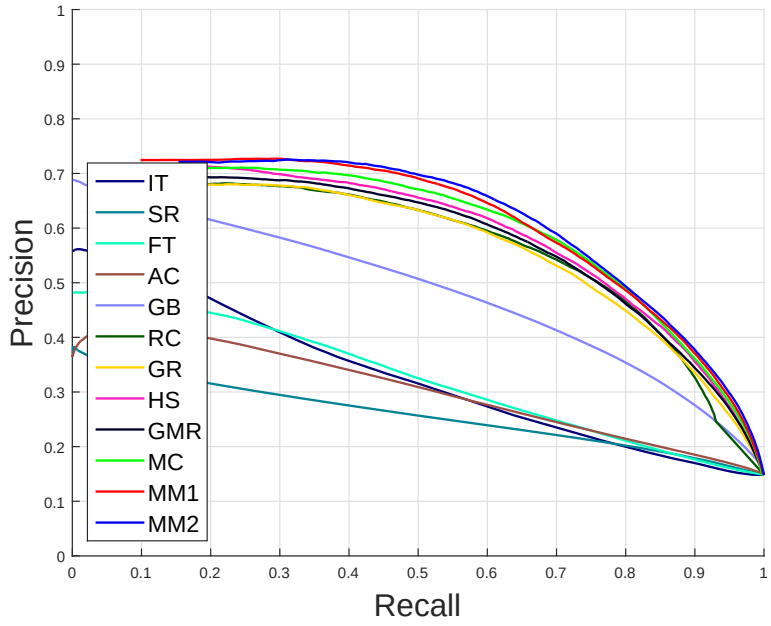


Figure 3.12: Visual comparison on ASD (1∼3 columns) and MSRA10K (4∼6 columns) datasets (*from top to bottom*: input images, RC, GR, HS, GMR, MC, proposed method (MM1), proposed method (MM2), and ground truth images).

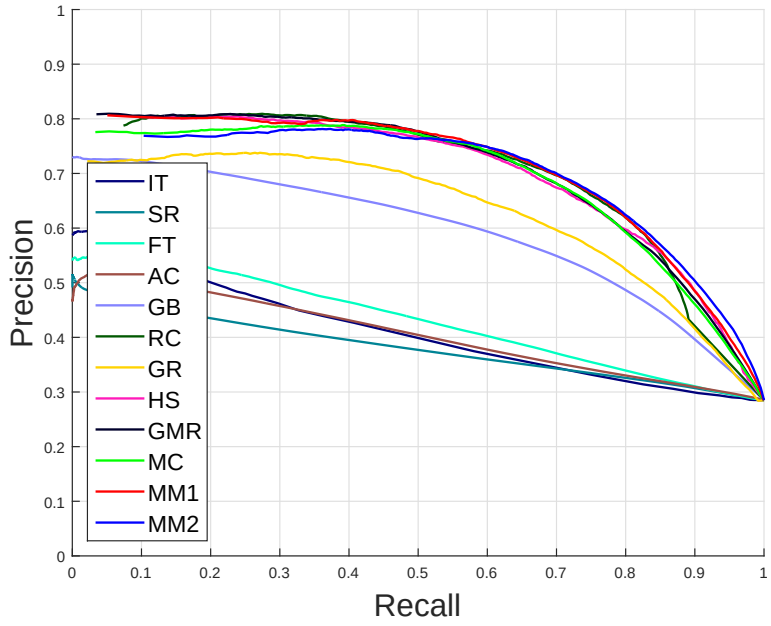
Complex datasets

Experiments are carried out on DUT-OMRON and ECSSD datasets which are the relatively complex datasets. As saliency detection has achieved great progress and reached high level performance in existing datasets, Yang *et al.* [11] created DUT-OMRON dataset for the purpose of saliency detection and eye fixation prediction to evaluate in more challenging and practical situation. The set consists of 5,168 images along with pixel-wise annotations which have one or more salient objects in each image with relatively complex background. Yan *et al.* [19] also constructed a complex dataset, i.e., ECSSD dataset for the same reason which consists of 1,000 images with accurately annotated images.

As expected, overall performance on DUT-OMRON and ECCSD datasets is decreased by a lot degree for all methods compared to ASD and MSRA10K datasets as shown in PR curves and ROC curves (Figure 3.13 and Figure 3.14). Though, it is also shown from these figures that the proposed methods detect salient objects more accurately than other methods. It demonstrates that the proposed methods are also competitive for the complex dataset as well as simple one. Unlike the two simple datasets, the proposed method based on color regularization shows better performance in terms of PR curve compared to the proposed spatial regularization method in entire ranges on DUT-OMRON dataset and high recall ranges on ECSSD. In case the DUT-OMRON and ECSSD sets, there are one or more salient objects which are possibly located at a distance, so the spatial regularization method may fail to detect some of these objects as demonstrated in 3rd column of Figure 3.16. Other algorithms also have trouble in localizing salient objects in this situation, however, the color regularization framework is capable of detecting multiple objects

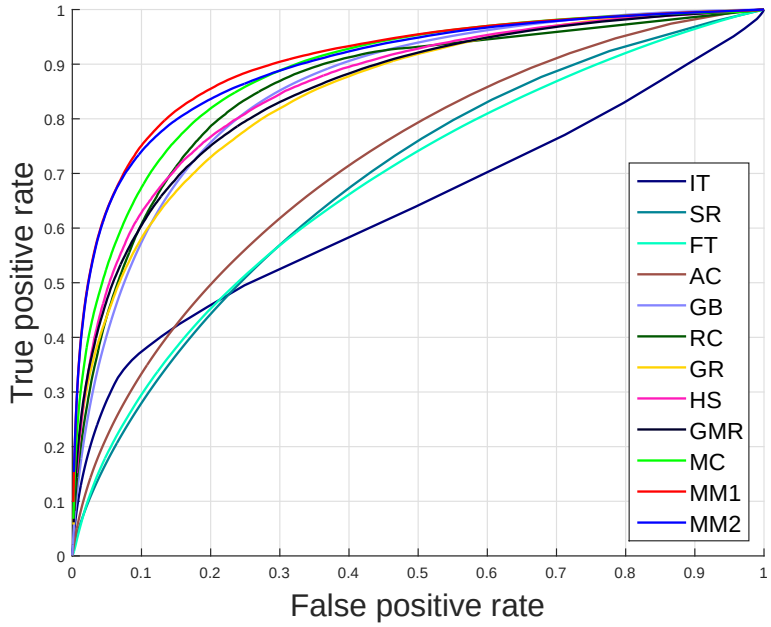


(a)

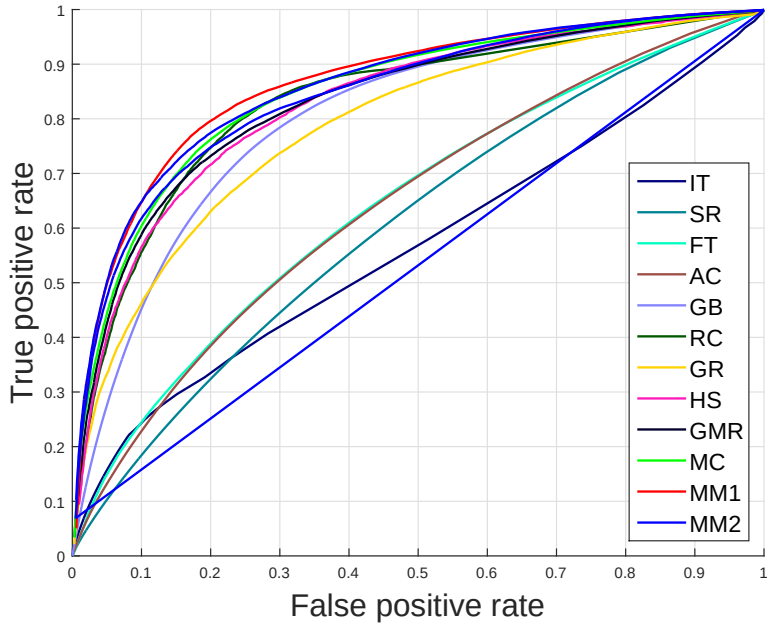


(b)

Figure 3.13: Comparison of PR curves on (a) DUT-OMRON and (b) ECSSD datasets.



(a)



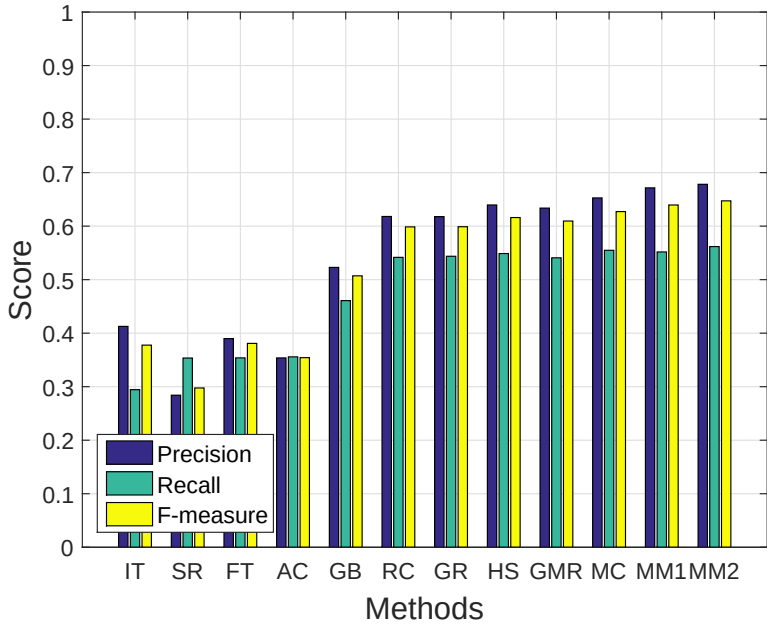
(b)

Figure 3.14: Comparison of ROC curves on (a) DUT-OMRON and (b) ECSSD datasets.

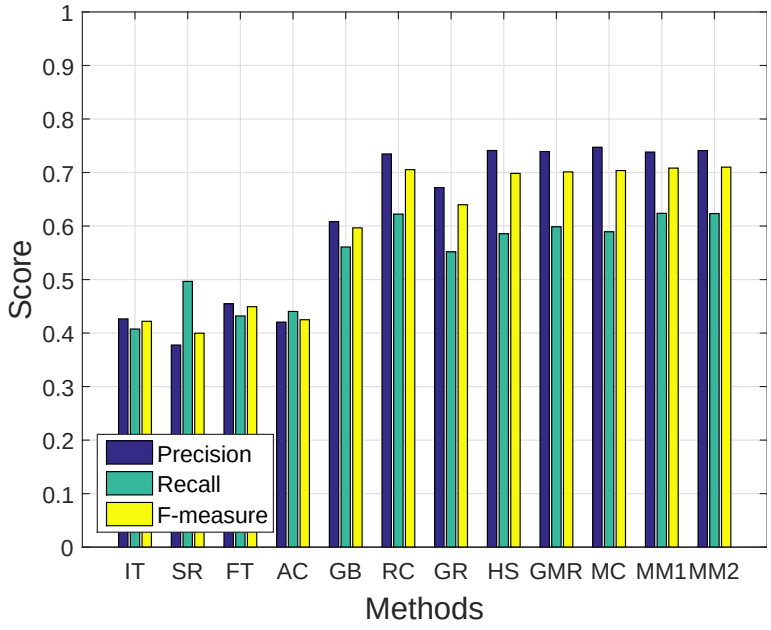
in spite of spatial distance because it can deal with multiple salient objects by means of propagation via the cluster layer. So, the color regularization method can be useful in more practical and general purposes than the spatial regularization method as well as the other methods.

It is also seen from Figure 3.15a that the color regularization method generates binarized saliency maps more accurately in terms of all objective measure, i.e., precision, recall and F-measure while the spatial regularization method places second in terms of precision and F-measure. Meanwhile, F-measures for the proposed methods are a little higher than other methods for ECSSD as shown in Figure 3.15b.

The color regularization has the merit of localizing multiple objects and complex shape objects such as 3∼5th columns of Figure 3.16, while the spatial regularization method is able to highlight salient objects in case of a single object with simple shape e.g., 6th column of Figure 3.16 and ASD and MSRA10K datasets. Both proposed methods are successful in suppressing background in a certain extent, however, these sometimes fail to suppress the background when salient objects are adjacent to complex background and also shares similar colors with them as the case of "Crane" in 2nd column of Figure 3.16, which is a common problem of adopting the regularization framework.



(a)



(b)

Figure 3.15: Comparison of precision, recall and F-measure at peak F-measure on (a) DUT-OMRON and (b) ECSSD datasets.

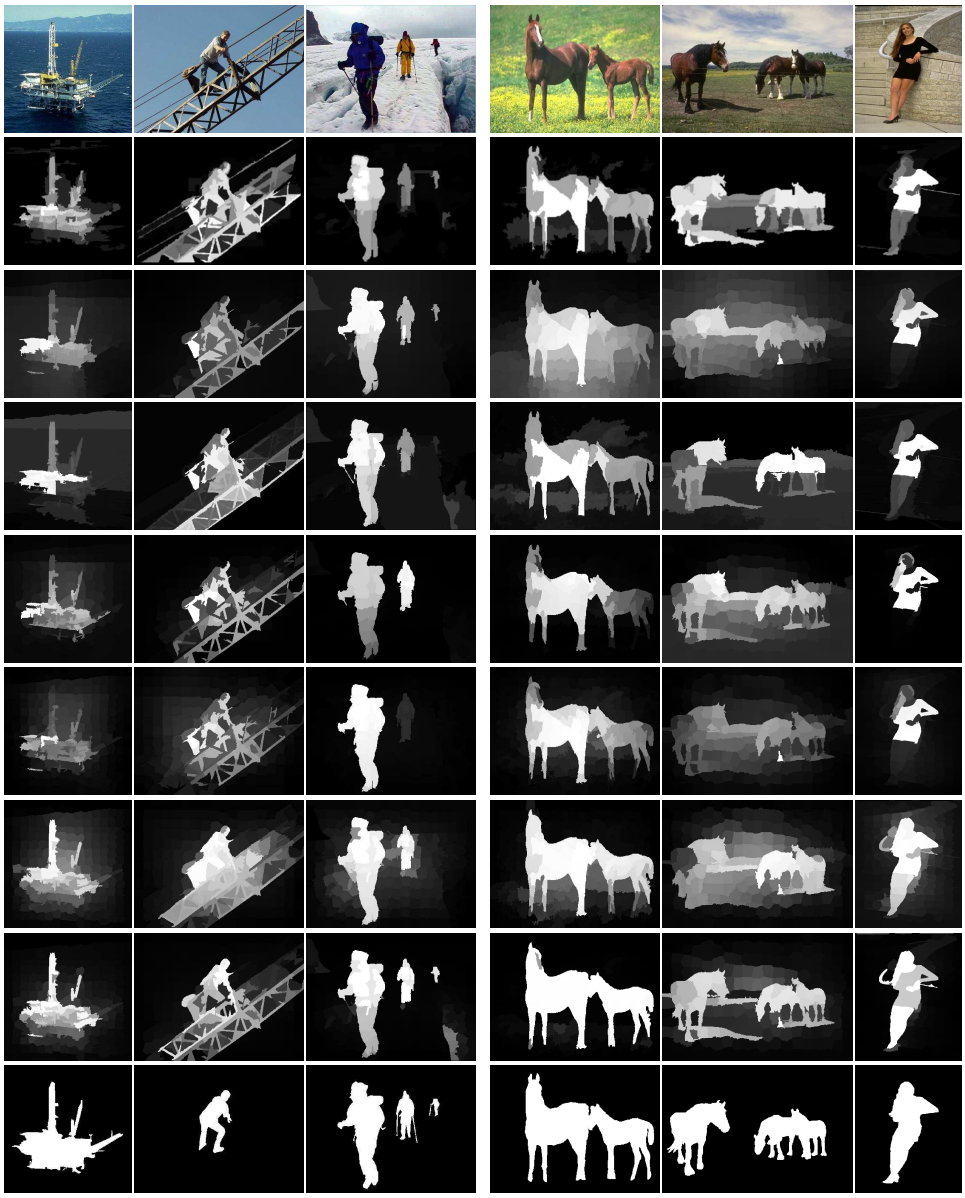


Figure 3.16: Visual comparison on DUT-OMRON (1∼3 columns) and ECSSD (4∼6 columns) datasets (*from top to bottom*: input images, RC, GR, HS, GMR, MC, proposed method (MM1), proposed method (MM2), and ground truth images).

Table 3.2: Comparison of AUCs-PR on four datasets.

	IT	**SR**	**FT**	**AC**	**GB**	**RC**
ASD	0.4526	0.3688	0.7060	0.5308	0.7060	0.9224
MSRA10K	0.4590	0.4549	0.6290	0.4996	0.7122	0.8707
DUTOMRON	0.3315	0.2586	0.3262	0.3035	0.4837	0.5656
ECSSD	0.4115	0.3803	0.4314	0.4052	0.5931	0.7016
	GR	**HS**	**GMR**	**MC**	**MM1**	**MM2**
ASD	0.9069	0.9263	0.9337	0.9315	**0.9459**	0.9327
MSRA10K	0.8415	0.8759	0.8801	0.8703	**0.8972**	0.8902
DUTOMRON	0.5674	0.5923	0.5804	0.6012	0.6126	**0.6176**
ECSSD	0.6306	0.7006	0.7039	0.6922	**0.7091**	0.6996

Table 3.3: Comparison of AUCs-ROC on four datasets.

	IT	**SR**	**FT**	**AC**	**GB**	**RC**
ASD	0.6303	0.6742	0.8647	0.8071	0.8647	0.9737
MSRA10K	0.6407	0.7356	0.7897	0.7563	0.9020	0.9358
DUTOMRON	0.6359	0.6881	0.6821	0.7213	0.8565	0.8593
ECSSD	0.5650	0.6071	0.6436	0.6425	0.8039	0.8320
	GR	**HS**	**GMR**	**MC**	**MM1**	**MM2**
ASD	0.9647	0.9688	0.9740	0.9778	**0.9865**	0.9793
MSRA10K	0.9253	0.9326	0.9435	0.9507	**0.9692**	0.9624
DUTOMRON	0.8464	0.8607	0.8528	0.8869	**0.9055**	0.8989
ECSSD	0.7874	0.8290	0.8337	0.8488	**0.8643**	0.8593

3.4 Limitations

Figure 3.17: Limitations of the proposed approach (*from left to right*: input image, result of MM1, result of MM2, and ground truth image).

Although experimental results show that the proposed method performs better than the others on average, there are some failure cases due to the limitations of contrast prior and regularization. First, background regions with high contrast often degrades the performance, because proposed saliency seed extraction method basically relies on the contrast. For example, Figure 3.17 shows the cases where background regions are so colorful that the foreground object is not well detected (top left figures). Second, when small regions within an object have very high saliency values, they make the other parts of the object to have comparably less saliency values (top right figures). In addition, non-saliency region between the salient objects sometimes appears to be salient due to the regularization framework of MM1 as in bottom left figures, while some parts of the salient object are not detected in case of MM2 as shown in bottom right figures when the parts are quite different in the color space.

Chapter 4

Co-saliency Detection for Multiple Images

4.1 Proposed Approach for Co-saliency Detection

The proposed approach for co-saliency detection is presented in this section. As the saliency approach, all images are represented by a unified multilayer graph based on color affinity. Intra image nodes are connected with its neighbors while inter image nodes are connected via cluster nodes. Saliency seeds are extracted using a region coherence cue as well as saliency cues, and background seeds are selected from boundary prior. Both seeds are propagated through the graph using a semi-supervised learning framework [12] and the co-saliency of each node is computed from the propagated results as the saliency case. Experiment results are presented with both analyses of the proposed approach and comparison of existing methods.

4.1.1 Graph Construction

The proposed method for detecting co-saliency is based on a graph that represents relations of intra image nodes as in the saliency detection method, and also the connections among the nodes of different images. In the proposed method, the graph has two types of nodes, i.e., image nodes and cluster nodes. A set of images, $\{I^m\}_{m=1}^{M}$, are represented by the image nodes where M is the number of images, and the cluster nodes are used to involve the image nodes of all images to a unified graph.

To get the image nodes and construct intra image graphs, each image in the set is independently segmented into over-segmented regions using the SLIC algorithm [64] at first. The over-segmented regions in the m-th image correspond to image nodes which are denoted as $\{v_i^m\}_{i=1}^{N_m}$, where v_i^m and N_m are the i-th image node and the number of image nodes in m-th image, respectively. An edge between the node i and j in the m-th image connects the pair of nodes with weight, w_{ij}^m which represents the affinity between nodes. In intra image graphs, the affinity is defined using color similarity which is identical to w_{ij} in equation (3.1). The nodes are sparsely connected to only their neighbor nodes which shares common boundaries, and the affinity matrix for the intra image graph of the m-th image is constructed whose (i, j)-th element is the weight between i-th and j-th node, defined as:

$$(\mathbf{W}^m)_{i,j} = \begin{cases} w_{ij}^m, & \text{if } j \in Q_i^m, \\ 0, & \text{otherwise,} \end{cases} \quad m = 1, \ldots, M \tag{4.1}$$

where Q_i^m is a set of neighbor nodes of node i.

Because $\mathbf{W}^m$ encodes relations of only intra image nodes, it is necessary to connect the nodes of different images for cooperating all the nodes together. In [37],

a graph matching method finds the matched pairs between images, and the matched pairs are connected each other with a matched score. Though ensuring good matched pairs for the images of similar scenes such as sequential frames of a video which are not severely different from each other, this approach easily fails to find those pairs in general, e.g., various background or different size of objects among images.

An indirect approach is introduced to overcome the problem in the proposed method. This approach basically ignores the connectivity between images, which means that there are no edges between inter image nodes. Thus, the graph of image nodes is represented as a form of block-wise diagonal matrix:

$$\mathbf{W}_I = \begin{bmatrix} \mathbf{W}^1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{W}^M \end{bmatrix}. \tag{4.2}$$

Instead, the proposed method introduces the cluster layer to consider the node relation between images. The cluster layer plays an important role of connecting the nodes of all images. In other words, the image nodes of different images are indirectly connected to one another via the nodes of the cluster layer.

To define the cluster nodes, K-means clustering is conducted using features of image nodes which are CIE Lab color vectors of corresponding nodes. It is noted that the feature $\mathbf{x}_i^m$ is the average color vector of all pixels belonging to image node i in the m-th image as in section 3.1.1. From the clustering, K clusters and its centroids, $\{C_i\}_{i=1}^K$, $\{\mathbf{c}_i\}_{i=1}^K$, are generated which are a set of cluster nodes and its average values on feature vectors belonging to corresponding clusters, respectively.

For connection between image nodes and cluster nodes, edges and their connec-

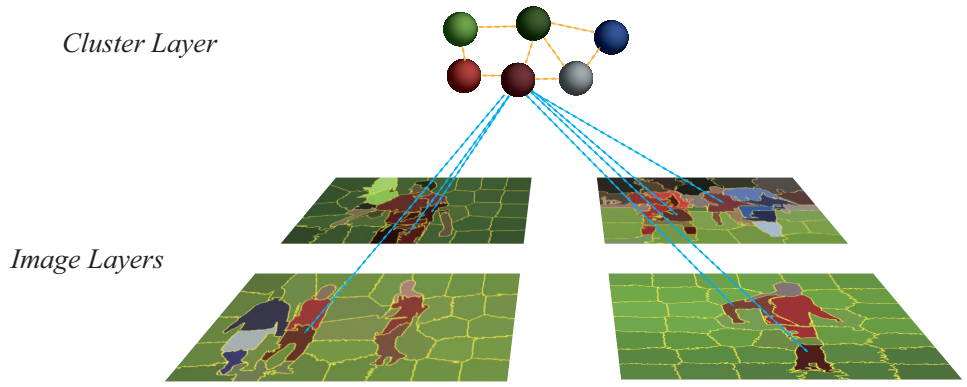


Figure 4.1: An illustration of the inter layer and intra cluster layer connections for multiple images (*blue dot lines*: inter layer edges between a cluster layer node and its corresponding nodes of the image layers, *yellow dot lines*: intra cluster layer edges).

tion weight are defined using color similarities between nodes as below:

$$w_{ij}^{IC} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{c}_j\|_2}{\sigma_c}\right), \tag{4.3}$$

where σ_c is a color control parameter for inter cluster similarity. Each image node is connected to only one cluster node that the image node belongs to, and it can be represented by an affinity matrix $\mathbf{W}_{IC}$ whose (i,j)-th element is written as:

$$(\mathbf{W}_{IC})_{i,j} = \begin{cases} w_{ij}^{IC}, & \text{if } i \in C_j, \\ 0, & \text{otherwise.} \end{cases} \tag{4.4}$$

Similar to the connection of intra image nodes, cluster nodes are connected to other nodes with color affinities. Here, a cluster node is connected to only k-nearest neighbors (k-NN), which means that the intra graph of cluster nodes is a sparse graph like the intra image graphs, which is written as:

$$w_{ij}^{C} = \exp\left(-\frac{\|\mathbf{c}_i - \mathbf{c}_j\|_2}{\sigma_c}\right), \tag{4.5}$$

$$\mathbf{W}_C = \begin{cases} w_{ij}^C, & \text{if } i \in k\text{-NN}(j) \text{ or } j \in k\text{-NN}(i), \\ 0, & \text{otherwise,} \end{cases} \tag{4.6}$$

where σ_c is a color control parameter for intra cluster similarity which is identical to the parameter used for inter cluster similarity. Finally, the overall graph for the proposed method is constructed from $\mathbf{W}_I$, $\mathbf{W}_{IC}$ and $\mathbf{W}_C$ which can be expressed as a block-wise matrix form:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_I & \mathbf{W}_{IC} \\ \mathbf{W}_{CI} & \mathbf{W}_C \end{bmatrix}, \tag{4.7}$$

where $\mathbf{W}_{CI}$ is the transpose of $\mathbf{W}_{IC}$.

4.1.2 Seed Extraction

In the saliency detection algorithm in chapter 3, it is essential to extract seeds for propagating them to all graph nodes. Similarly, seed extraction is also required to compute co-saliency because the philosophy of the algorithms is basically identical. In the saliency detection, saliency seeds are selected based on two priors, i.e., contrast prior and center prior. The contrast prior assumes that salient objects have distinctive features so that they show high contrast to other regions while the center prior is based on the observations that those objects are likely to be located in the image center. From these priors, the initial saliency (IS) of each image is computed, which has been defined as s_i in equation (3.6), and the nodes with high saliency values are selected as saliency seeds.

However, seed extraction of the saliency case is not appropriate for co-saliency detection because IS does not consider the correlation among salient objects so that some salient objects in a specific image can be detected. To overcome the problem,

similarity of saliency (SoS) is introduced, which basically estimates the similarities among salient objects to filter out individual saliency objects in the set of images.

To compute SoS, salient regions should be defined in advance. IS is binarized to 1 or 0 for salient region and non-salient region respectively as:

$$b_i^m = \begin{cases} 1, & s_i^m \geq \tau_m, \\ 0, & s_i^m < \tau_m, \end{cases} \tag{4.8}$$

$$\tau_m = \max\left(\frac{1}{N_m}\sum_{i=1}^{N_m} s_i^m, \quad 0.5\right), \tag{4.9}$$

where τ_m is an image adaptive threshold for the m-th image defined as the average value of the IS, and the saliency regions of m-th image correspond to B_m. Since co-saliency shares common salient objects among images, co-saliency nodes are expected to have high similarity to B_m. From this, SoS for a node is defined as the sum of similarity to $\{B_m\}_{m=1}^M$ of all images, which is written as:

$$t_i^m = \sum_{p \neq m} \sum_{q \in B_p} \exp\left(-\frac{\|\mathbf{x}_i^m - \mathbf{x}_q^p\|}{\sigma_t}\right), \tag{4.10}$$

where σ_t is a similarity control parameter for SoS which penalizes the similarity more in case of smaller σ_t.

Though SoS itself can highlight co-salient objects in some degree, it is insufficient to suppress the background regions. Therefore, initial co-saliency (IC) for seed extraction is obtained by combining IS and SoS with multiplication operator as:

$$u_i^m = s_i^m \cdot t_i^m. \tag{4.11}$$

Similar to the case of the saliency detection, top $(100 \times \gamma)\,\%$ of co-saliency regions in each image are extracted as co-saliency seeds.

Meanwhile, the process for background seed extraction is identical to the case of the saliency detection, which selects boundary nodes of images as the background

Figure 4.2: Example images for co-saliency cues (*from top to bottom*: input images, initial saliency (IS), binarized saliency maps, similarity of saliency (SoS), and combined co-saliency result (IC) of the *Women Soccer Players* set).

seeds based on the boundary prior of saliency detection. In addition, seeds selected as both co-saliency and background seeds are excluded from both seeds because those seeds are not reliable. In summary, saliency seeds are defined as:

- Saliency seeds ($\mathbf{y}_{I,s}$) : high IS and SoS nodes which are not on the image boundary,

- Background seeds ($\mathbf{y}_{I,b}$) : low IS and SoS nodes on the image boundary.

4.1.3 Seed Propagation

From the co-saliency and background seeds, co-saliency is computed by propagating them to all nodes in the set of images. Similar to the propagation of saliency detection, a graph-based learning method is adopted for effective propagation [12], which makes the full pairwise graph as:

$$\mathbf{W}_L = (1-\alpha)\left(\mathbf{D} - \alpha\mathbf{W}\right)^{-1}, \tag{4.12}$$

where α is a learning balance parameter and $\mathbf{D} = diag\left\{d_1, d_2, \cdots, d_N\right\}$ is the degree matrix of $\mathbf{W}$ whose diagonal entry is the degree of each node ($d_i = \sum_j w_{ij}$). As mentioned in section 4.1.1, there is no connection between inter image nodes in the graph with affinity matrix $\mathbf{W}$, though inter image nodes are indirectly connected to each other. However, the learned graph, $\mathbf{W}_L$, has full pairwise relations of all nodes. In other words, the inter image nodes of the graph are directly connected so that graph ensures straightforward propagation between images.

For co-saliency detection, overall affinities to the co-saliency and background seeds are computed, which is written as:

$$\mathbf{f}_s = \mathbf{W}_L\ \mathbf{y}_s = \sum_{i \in S_s} \mathbf{w}_L^i, \tag{4.13}$$

$$\mathbf{f}_b = \mathbf{W}_L\ \mathbf{y}_b = \sum_{i \in S_b} \mathbf{w}_L^i, \tag{4.14}$$

where $\mathbf{y}_s = [\mathbf{y}_{I,s}; \mathbf{0}]$, $\mathbf{y}_b = [\mathbf{y}_{I,b}; \mathbf{0}]$ are seed vectors of co-saliency and background which are concatenated with a zero vector for cluster nodes, and S_s and S_b represent saliency and background seed sets respectively. $\mathbf{f}_s$ and $\mathbf{f}_b$ are decomposed to node vectors of each image and cluster, $\mathbf{f}_s = \left[\mathbf{f}_s^1; \ldots; \mathbf{f}_s^M; \mathbf{f}_s^C\right]$ and $\mathbf{f}_b = \left[\mathbf{f}_b^1; \ldots; \mathbf{f}_b^M; \mathbf{f}_b^C\right]$, and finally, the co-saliency of m-th image is defined as:

$$\mathbf{z}^m = \left(\mathbf{f}_s^m - \mathbf{f}_b^m\right) ./ \left(\mathbf{f}_s^m + \mathbf{f}_b^m\right), \tag{4.15}$$

where ./ is the element-wise division of two vectors. The numerator represents the saliency while the denominator maintains the balance among nodes, and it is normalized to $[0, 1]$.

4.2 Experiments

In this section, experimental results of the proposed method are presented from three datasets of co-saliency detection: CPD [36, 70], iCoseg [71, 72] and iCosegSub [71, 72]. The experiments are conducted to choose the parameters used in the proposed method by varying them, and also analysis for the parameters is presented. In order to demonstrate benefits of the proposed method, extensive experiments are carried out and the results of the proposed method are compared to the state-of-the-art methods as well as the saliency detection method (MM2) in section 3. The comparison is conducted in terms of quantitative measures and visual comparison. For quantitative measures, precision-recall (PR) curve, ROC curve and area under curve (AUC) are used, and precision, recall and F-measure in a fixed threshold maximizing the F-measure are also provided to compare the performance of binarized co-saliency detection results all of which are identical to the measures employed in the saliency detection evaluation, referring to section 3.3.1. In addition, the visual comparison is shown in order to demonstrate how each algorithm works and to point out the difference among them.

4.2.1 Experiment Setup

The number of super-pixels is set to $N_m = 250$ for all images ($m = 1, \ldots, M$). It is noted that the number of images in a set M, varies depending on the set.

The number of the clusters and the nearest neighbors for the intra cluster nodes connection are set to $K = 50$ and $k = 5$ in all experiment regardless of the M. The σ_c is the color control parameter for cluster node affinity which is designed to 0.05 while the color control parameter for SoS, σ_t, is set to 0.4. In the case of the cluster node affinity, relatively low value of σ_c tends to connect the nodes sharing high similarity, which makes the nodes to be sparsely connected, and it is expected to learn more desired fully pair-wise graph. Meanwhile, σ_t is set to comparably high values so that it smoothly measures the similarity between nodes, otherwise it can cause mis-detection of saliency nodes when the colors among the images are quite different. As in the chapter 3, the α and γ are set to 0.99 and 0.1, respectively.

4.2.2 Analysis of the Proposed Approach

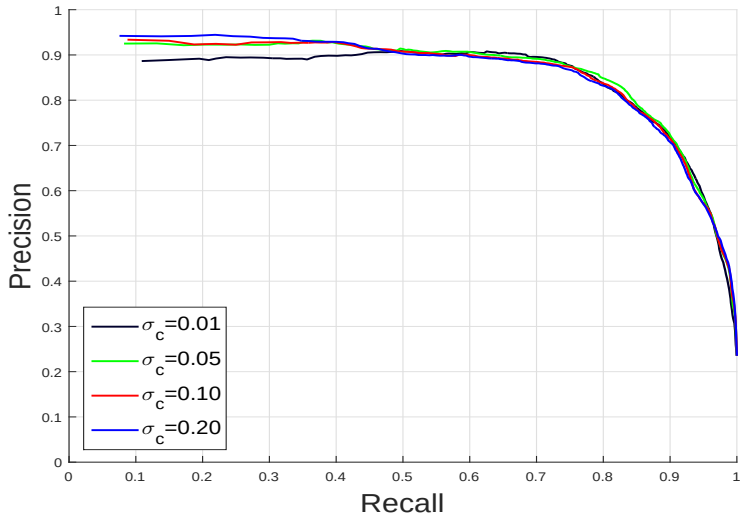
There are some parameters for the proposed method. In this subsection, the proposed method is investigated in the view of how the parameters work. In order to determine the parameters, precision-recall curves are plotted for some variation of the parameters using the iCosegSub dataset consisting of 5 images in a set which can be considered to be applied to real situation. Also, the visual comparison for each parameter is accompanied for easy explanation.

Effectiveness of the color control parameter for cluster affinity (σ_c)

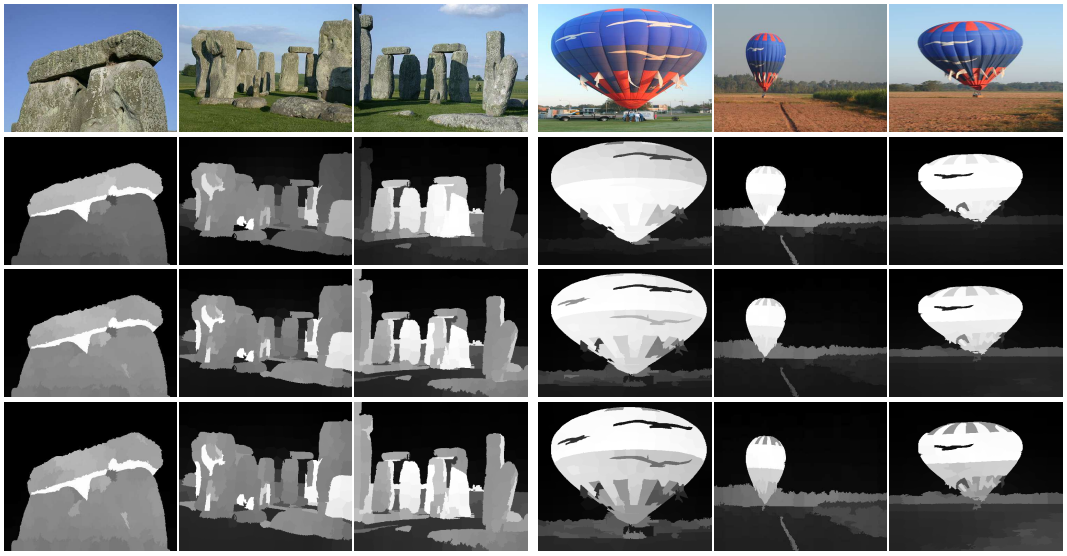
Firstly, the control parameter for cluster affinity σ_c, is examined here. It concerns strength of the graph among intra cluster nodes as well as between cluster nodes and image nodes. Contrast to the intra node connection which uses the adaptive color control parameter, a fixed value is applied to σ_c because the characteristics of nodes in various images and clusters are so different that it is hard to handle together in this

case. When σ_c is large, the weight in equations (4.3) and (4.5) becomes larger, and the connection between nodes is stronger. Meanwhile, the weight is smaller in the case of small σ_c so that connection with lower color similarity is strongly suppressed compared to the connection with high color similarity. Thus, the gap between the cases of small and large color affinity is become larger, and the connection of the graph tends to be more sparse.

The PR curves and co-saliency detection results obtained with various σ_c are shown in Figure 4.3. It can be seen that smaller σ_c yields better results in high recall ranges while yielding worse results in low recall ranges as shown in Figure 4.3a. It implies that the smaller σ_c case may fail to capture co-salient objects with high values, but it has the merit of reducing false positive errors as demonstrated in Figure 4.3b. In contrast to smaller σ_c case, it captures co-salient objects well while it also highlights background regions in a certain degree, where it is reflected in the PR curves which show better results in low recall ranges while showing worse results in high recall ranges. In the proposed method, σ_c is set to 0.05 because it is well balanced.



(a)

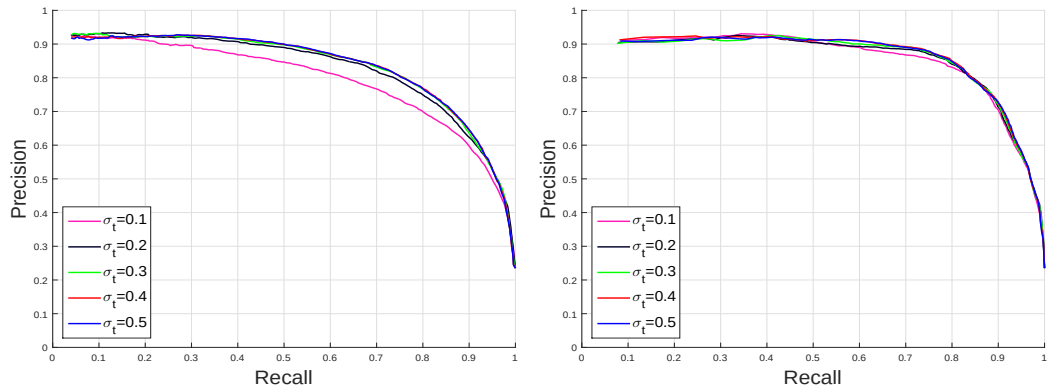


(b)

Figure 4.3: Comparison of results with different σ_c: (a) PR curves on iCosegSub set, and (b) visual comparison on *Stonehenge* and *Hot Balloons* sets (*from top to bottom*: input images, results with $\sigma_c = 0.01$, $\sigma_c = 0.05$, and $\sigma_c = 0.20$).

Effectiveness of the color control parameter for SoS (σ_t)

The parameter σ_t plays a role in controlling similarity values defined by the summation of similarities to salient objects. It is also an exponential form like the affinity weight of the graph so that it works similar to the case of σ_c. Therefore, smaller σ_t makes more gap of IC as well as SoS between high and low color distances so that background and individual salient objects are effectively suppressed as shown in the *Taj Mahal* set of Figure 4.4c where 2nd and 3rd rows of the figure are IC with $\sigma_t = 0.10$ and $\sigma_t = 0.40$ respectively. But, there is a trade-off that the small σ_t can yield false negative pixels when the colors are quite varied among a set of images due to certain conditions, such as illumination changes. When the color difference is so high, it can highlight common background regions more than co-salient regions demonstrated in the *Stonehenge* set of Figure 4.4c. It means that IC is more sensitive when small σ_t is adopted. Figure 4.4a shows PR curves for IC with different σ_t where the cases of larger σ_t have better results plotted upper in most ranges. It is also shown that co-saliency detection results from the different IC values make similar aspects as the IC results, which is demonstrated in Figure 4.4b. However, the gaps of the plotted graphs become smaller in the co-saliency results because the seed propagation step in the proposed method uses both of co-salient and boundary seeds, and the boundary seeds contributes to suppressing the background regions. Figure 4.4c shows how the seed propagation affects when the co-salient seeds are extracted from different IC according to σ_t. In the proposed method, σ_t is set to comparably larger value, 0.40, for the above reason. It is noted that too large σ_t makes SoS uniform which does not contribute to co-saliency detection anymore.



(a) (b)

(c)

Figure 4.4: Comparison of results with different σ_t: (a) PR curves of IC on iCosegSub set, (b) PR curves of final results on iCosegSub set, and (c) visual comparison on *Stonehenge* and *Taj Mahal* sets (*from top to bottom*: input images, IC with $\sigma_c = 0.1$, IC with $\sigma_c = 0.4$, final results with $\sigma_c = 0.1$, and final results with $\sigma_c = 0.4$).

4.2.3 Comparison with Other Algorithms

Experiments are conducted on the CPD, iCoseg and iCosegSub datasets. The results and discussions will be individually represented according to the datasets. The proposed method is compared with some of the state-of-the-art methods: CP [36], CB [41], HS [43], MG [46]. CB and MG results have been obtained from the author's codes provided in web sites. It is noted that MM2 is adopted for intra saliency maps of the MG since the code requires the intra saliency of their previous work which is not provided. Author's results of the HS on the CPD and iCoseg datasets have been downloaded from a web. CP results are only provided for CPD dataset because it is only able to deal with a pair of images. In addition, the single image saliency detection method based on the cluster layer (MM2) in chapter 3 is included in the comparison. As the approach of the saliency detection is basically same to that of the proposed co-saliency detection method, comparison with MM2 is very effective to verify influences of the co-saliency cue and how the saliency and co-saliency detection methods differently work.

Co-saliency Pairs Dataset

The co-saliency pairs dataset (CPD) is rooted in a co-segmentation dataset containing contains 25 pairs of images [70], and Li *et al.* [36] brought the dataset in co-saliency research and added another 85 pairs of images. Therefore, CPD consists of 105 image pairs with hand-labeled ground truth. As CPD dataset contains a single object for each image, mostly located near the image center, it is considered to be relatively easy dataset.

Figure 4.5 compares PR curves of various methods. The proposed method outperforms the CP and CB on a entire recall range and shows slightly better results than the HS, MG methods, which means that the proposed method precisely detects co-salient regions in a given recall constraint. The single image saliency detection method MM2 shows comparable or better results in high recall ranges since it highlights salient objects more uniformly and consequently detects more true positive pixels in a low threshold as shown in Figure 4.8 (e.g., *Mascot* set). However, it has poor performance in low and medium recall ranges because of detecting some background regions with high values. ROC curves in Figure 4.6 show that HS, MG, MM2 and proposed method yield comparable results while curves for CP and CB are located below the others in a certain degree. In comparison with the PR curves, high recall ranges (high TPR) are so prominent in the ROC curves that MM2 is slightly better than the proposed method as well as other methods, which is also demonstrated in AUCs in Table 4.1 and 4.2. Because saliency objects mostly correspond to co-saliency objects and the number of image sets is restricted to two in CPD dataset, proposed method has a little advantage over MM2.

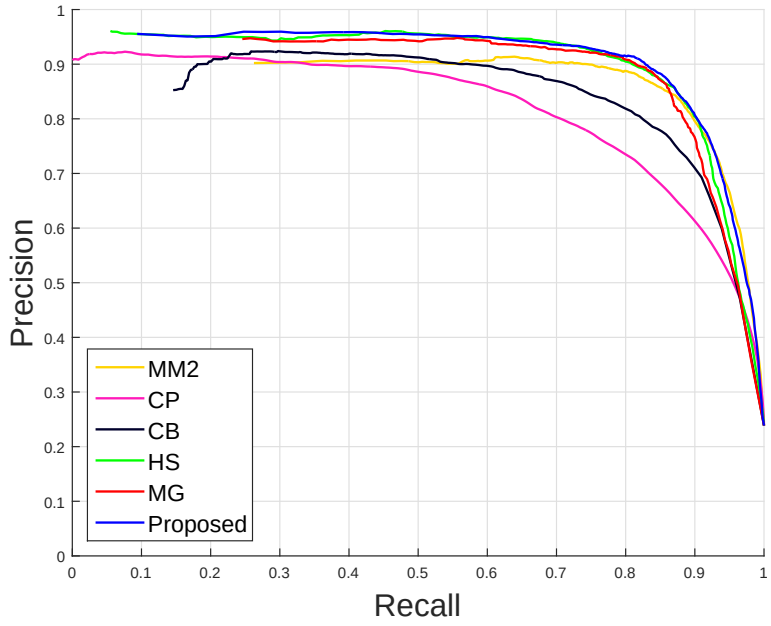


Figure 4.5: Comparison of PR curves on CPD dataset.

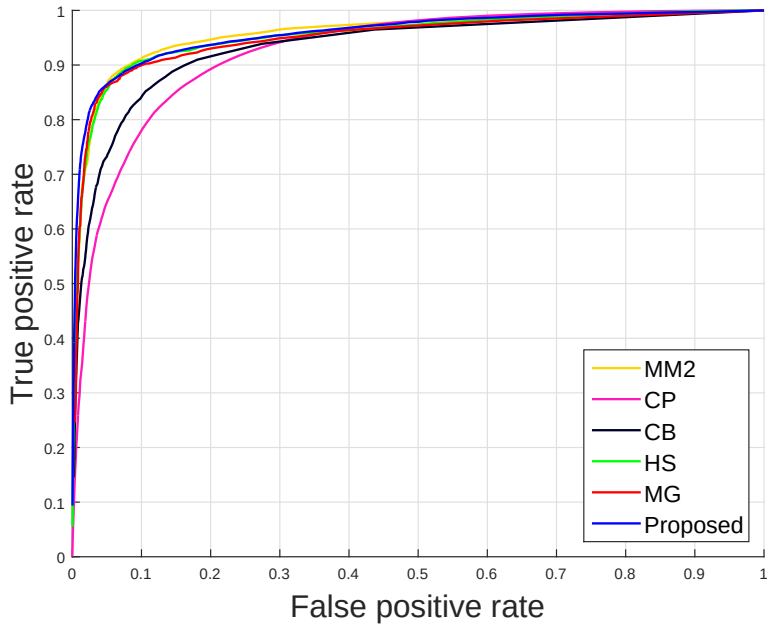


Figure 4.6: Comparison of ROC curves on CPD dataset.

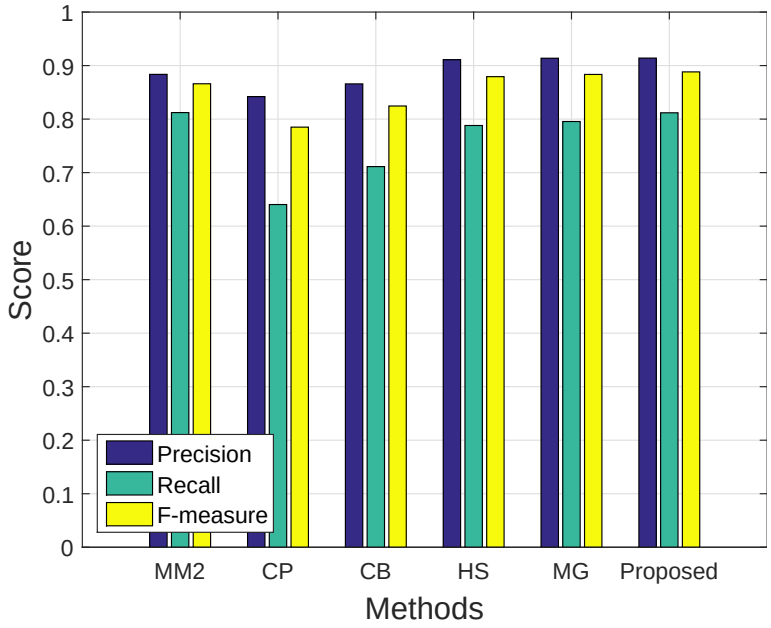


Figure 4.7: Comparison of precision, recall and F-measure at peak F-measure on CPD dataset.

Similar to above results, it can be seen from Figure 4.7 that HS, MG, MM2 and the proposed method are comparable in precision, recall and F-measure. Figure 4.8 shows the visual comparison of the proposed method with all other methods. CP fails to capture co-salient objects such as *Mascot* and *Sheep* image sets, while CB and HS cannot highlight co-salient objects uniformly for all example images, because it does not consider local consistency of pixels or segmented regions. MM2 overcomes those those problems in most images, but it sometimes fails to detect co-salient objects when there are isolated regions of very high contrast as in *Train* set. Though MG and the proposed method also undergo the non-homogeneous detection problem, e.g, eyes in the *Mascot* set, these methods detect most of co-salient regions with uniformly high value.

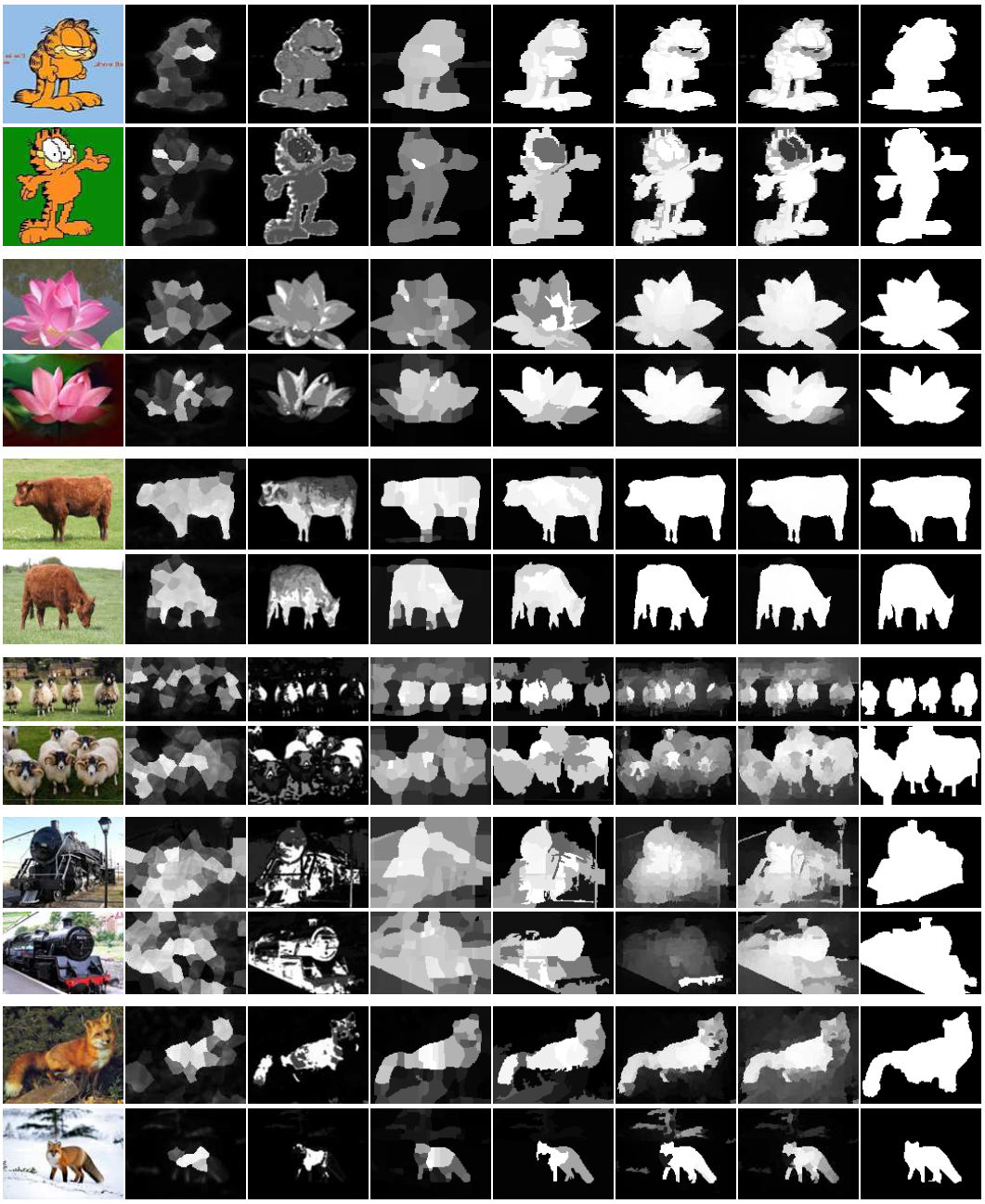


Figure 4.8: Visual comparison on CPD dataset (*from left to right*: input images, CP, CB, HS, MG, MM2, proposed method, ground truth images, *from top to bottom*: *Mascot*, *Flower*, *Cow*, *Sheep*, *Trains*, *Fox* image pairs).

iCoseg Dataset

The CMU-Cornell iCoseg dataset [71, 72] is employed for the evaluation of the proposed method compared with other methods. It was created for the purposed of co-segmentation research at first like CPD set, which is widely exploited for co-saliency research as well. The iCoseg consists of 38 images sets and totally 463 images with pixel-wise ground truth images, where each set contains 4 to 41 images. Though co-salient objects are mostly located in the image center, there are comparably a lot of sets containing co-salient objects adjacent to image boundary in the iCoseg dataset.

The proposed method outperforms other methods on extensive ranges of recall as shown in Figure 4.9, while HS and MG yield a little better performance on the ranges of over 0.9 in recall. For ROC curves plotted in Figure 4.10, ROC curve of the proposed method is higher on small false positive rate and HS and MG is higher on the ranges between 0.1 and 0.35 which correspond to the ranges that those algorithm show better in the PR curves. It is noted that a small false positive rate is generally more desirable to be applied to other applications. Unlike the case of the CPD dataset, both PR and ROC curves for MM2 are lower than the proposed method, which can be analyzed in twofold: there are individual salient objects in the iCoseg dataset, and each set contains so many images that co-salient algorithms can exploit extensive information from those images in contrast to the single image saliency algorithm. In a result, the AUCs for PR and ROC curves in Table 4.1 and 4.2 demonstrate that the proposed method comprehensively performs better.

An additional experiment is conducted to evaluate co-saliency maps when a threshold is applied to generate binary saliency maps. It is shown from Figure 4.11

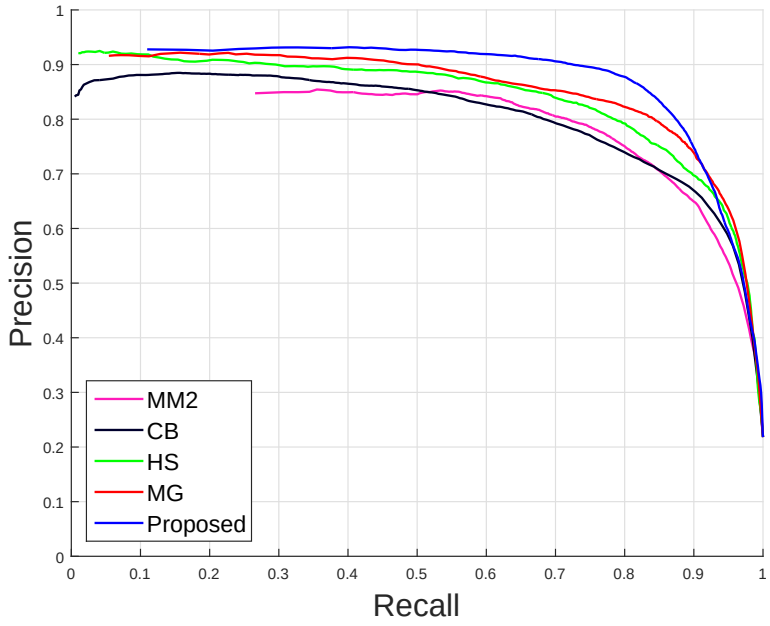


Figure 4.9: Comparison of PR curves on iCoseg dataset.

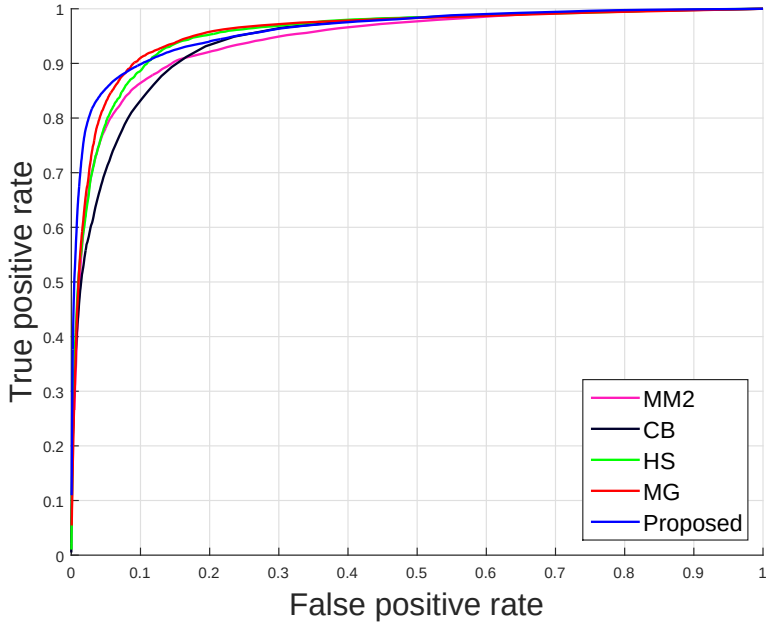


Figure 4.10: Comparison of ROC curves on iCoseg dataset.

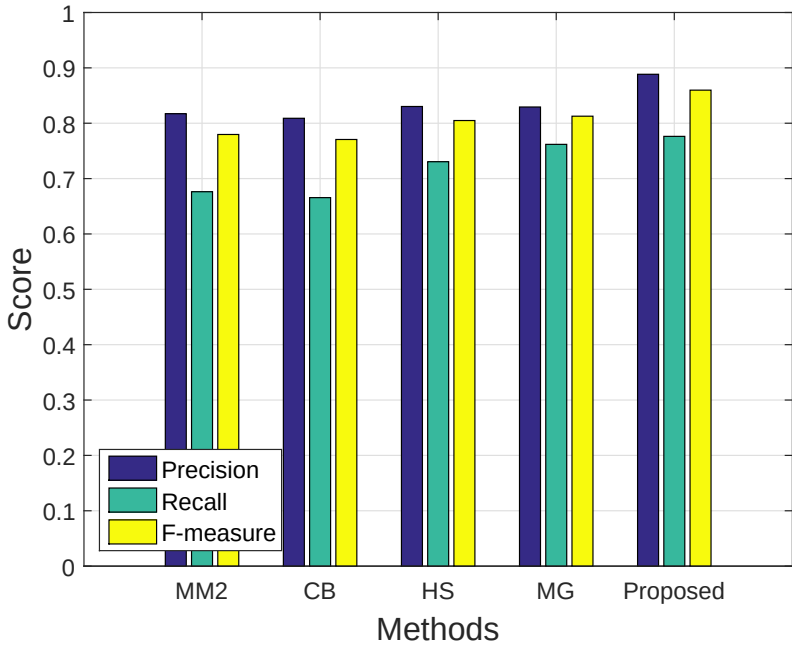


Figure 4.11: Comparison of precision, recall and F-measure at peak F-measure on iCoseg dataset.

that proposed method outperforms other methods in terms of all of quantitative measures i.e., precision, recall and F-measure, which demonstrates that proposed method has the merit of practical use for extensive applications.

Figure 4.12 and 4.13 show some example images of results for the intuitive comparison, where CB fails to highlight co-salient objects in general while HS suppresses co-salient objects around image boundaries such as *Duck* and *Cheetah* sets because the process of HS is highly depending on the boundary adjacency prior. MG and MM2 can capture saliency objects well, but they are sometimes unsuccessful in filtering out individual salient objects, e.g., bronze-winged duck in the *Duck* set and red players in the *Women Soccer Players* set. The proposed method overcomes these problems though some individual salient objects are not perfectly suppressed.

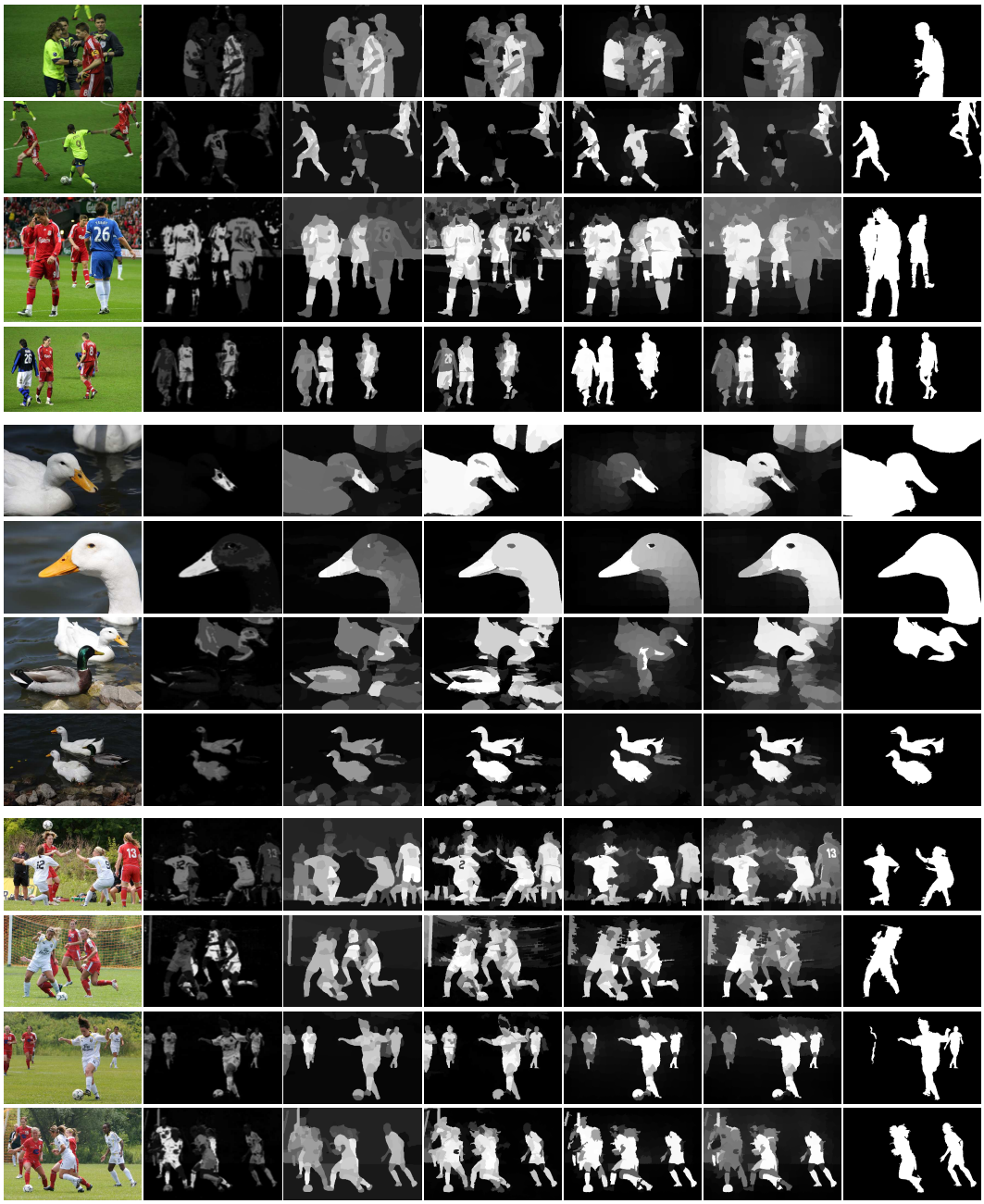


Figure 4.12: Visual comparison on iCoseg dataset (*from left to right*: input images, CB, HS, MG, MM2, proposed method, ground truth images, *from top to bottom*: *Liverpool*, *Duck*, *Woman Soccer Players* sets).

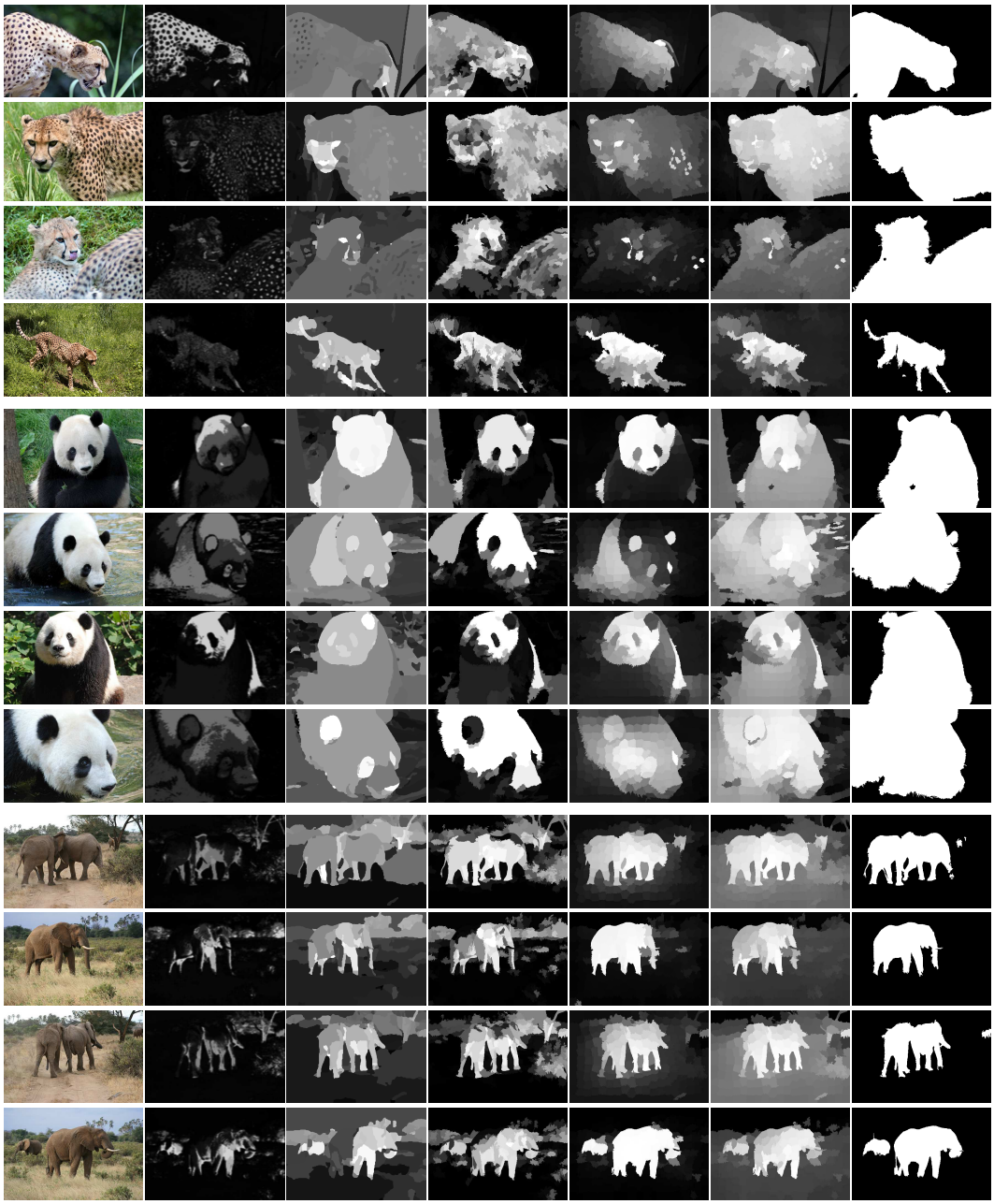


Figure 4.13: Visual comparison on iCoseg dataset (*from left to right*: input images, CB, HS, MG, MM2, proposed method, ground truth images, *from top to bottom*: *Cheetah*, *Panda*, *Elephant* sets).

iCosegSub Dataset

As a subset of the iCoseg set, the iCosegSub dataset [71, 72] is composed of 37 sets, each of which contains equally 5 images sampled from the iCoseg dataset. Though it was designed to keep statistical comparable across sets in co-segmentation evaluation, it is also meaningful for co-saliency detection in the way that it is more applicable as well as enough to take advantage of multiple images. Figure 4.15 and 4.16 are PR curves and ROC curves of various methods respectively which show similar tendency to the case of iCoseg dataset because of redundancy between two sets. So, the proposed method is mostly higher than other methods on PR curves while showing better results in low FPR ranges on ROC curves as the iCoseg dataset case.

In comparison with iCoseg dataset, it can be also seen from Figure 4.14 that the results on the iCosegSub are worse than those on the iCoseg. The reason is that co-saliency is computed from only 5 images on iCosegSub, but 18 images on iCoseg, which is more sufficient to compute the co-saliency cue.

Figure 4.14: Comparison of co-saliency detection results using different number of images (*from top to bottom*: input images, results computed from 5 images (iCosegSub) and results computed from 18 images (iCoseg)).

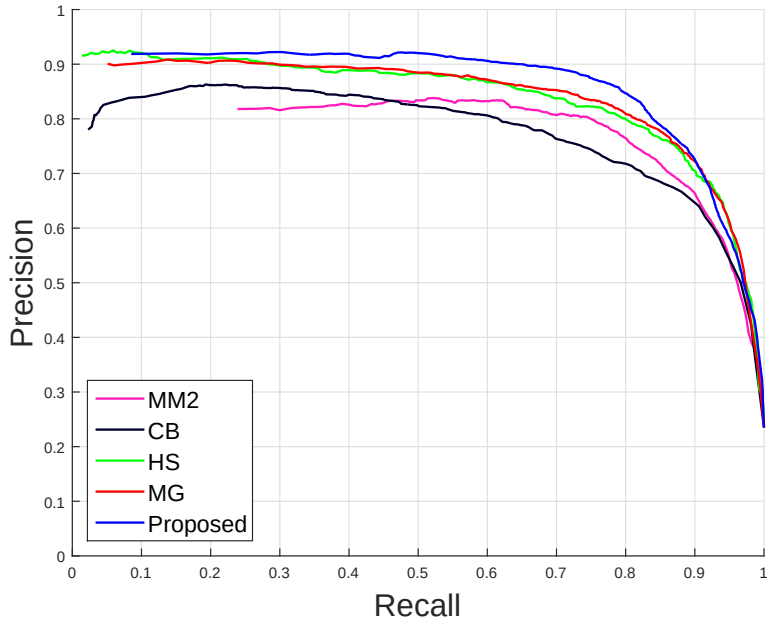


Figure 4.15: Comparison of PR curves on iCosegSub dataset.

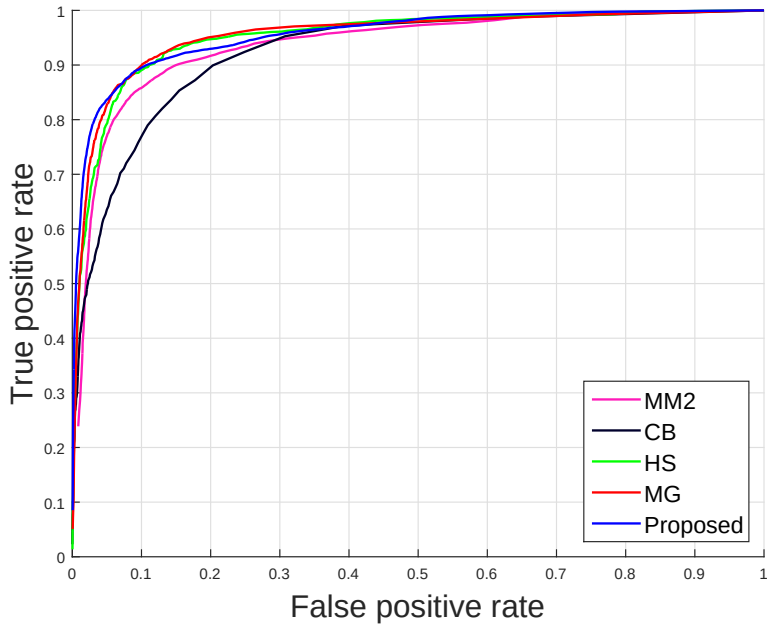


Figure 4.16: Comparison of ROC curves on iCosegSub dataset.

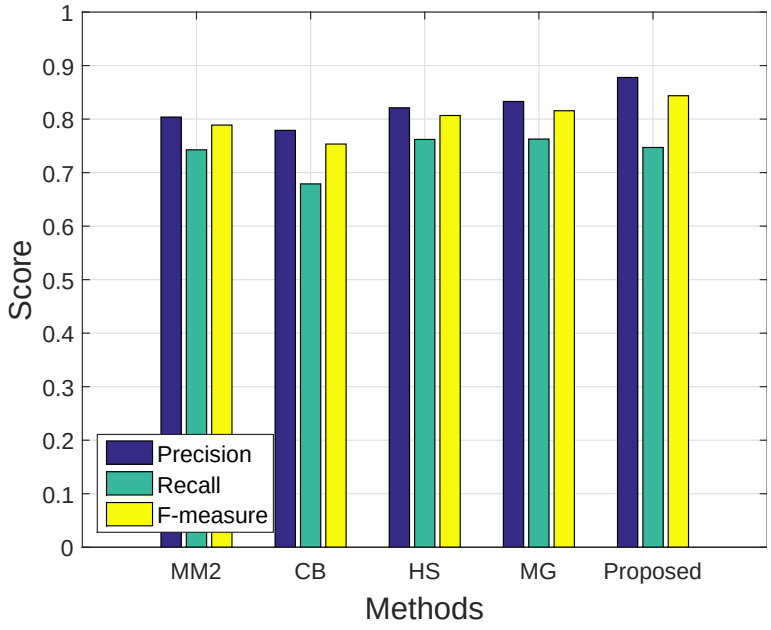


Figure 4.17: Comparison of precision, recall and F-measure at peak F-measure on iCosegSub dataset.

Figure 4.17 plots precision, recall and F-measure of co-saliency maps binarized with a threshold. Precision of the proposed method is a lot better than other methods while recall of the proposed method is comparable to others, which consequently leads F-measure of the proposed method to the best among all methods.

In visual comparison, it is shown from Figure 4.18 that the proposed method highlights co-salient objects consistently in general, except the first image of the *Red Sox Players* set where the player on the image wears a protector of red color which is uncommon in other images. Meanwhile, other methods fail to uniformly detect co-saliency objects in spite of same objects among images, i.e., *Christ the Redeemer* set.

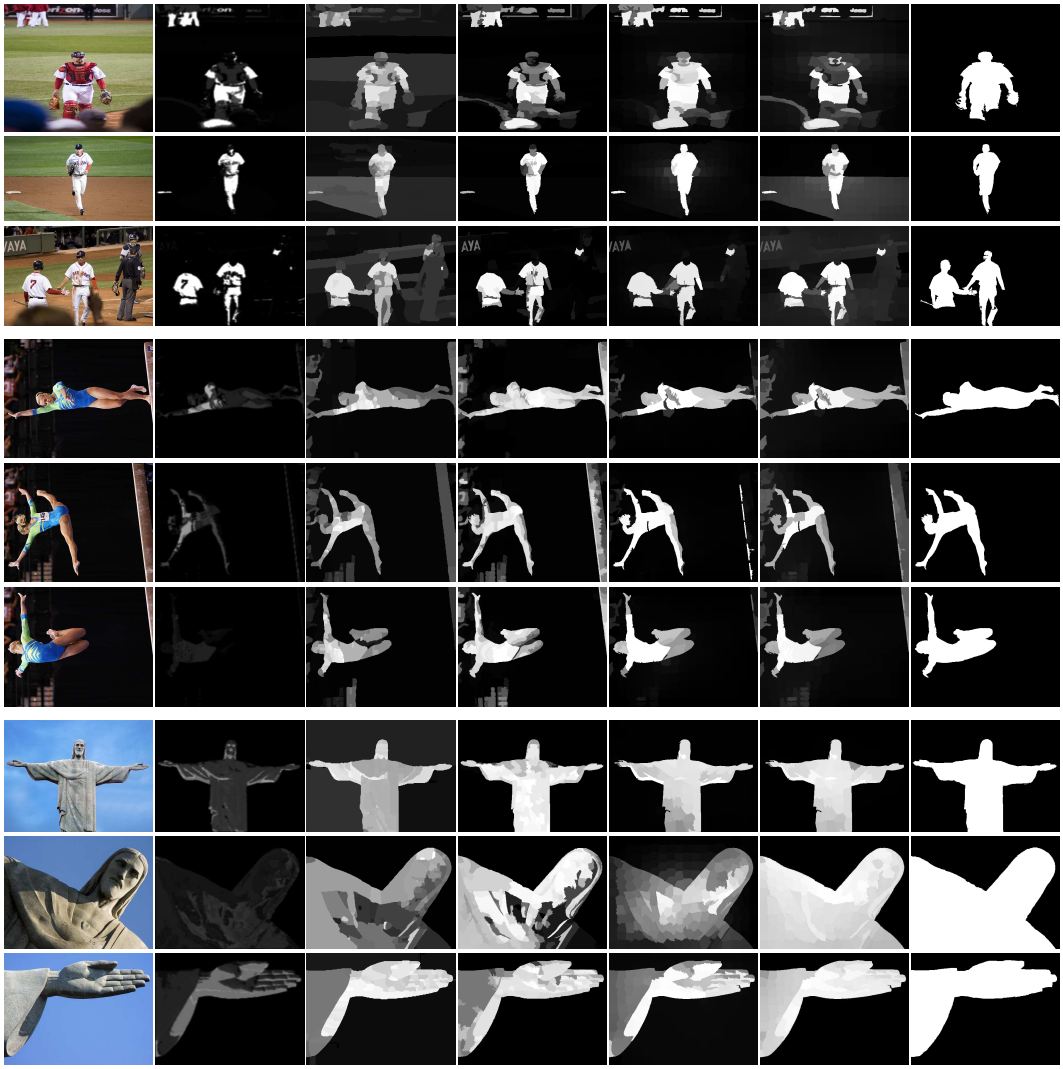


Figure 4.18: Visual comparison on iCosegSub dataset (*from left to right*: input images, CB, HS, MG, MM2, proposed method and ground truth images, *from top to bottom*: *Bear*, *Red Sox Players*, *Gymnastics*, *Christ the Redeemer* sets).

Table 4.1: Comparison of AUCs-PR on three datasets.

	CPD	iCoseg	iCosegSub
MM2	0.8702	0.7911	0.7788
CP	0.8190	N/A	N/A
CB	0.8402	0.8035	0.7754
HS	0.9022	0.8393	0.8497
MG	0.8912	0.8541	0.8411
Proposed	**0.9089**	**0.8772**	**0.8633**

Table 4.2: Comparison of AUCs-ROC on three datasets.

	CPD	iCoseg	iCosegSub
MM2	**0.9588**	0.9435	0.9373
CP	0.9256	N/A	N/A
CB	0.9335	0.9423	0.9250
HS	0.9542	0.9546	0.9526
MG	0.9503	0.9574	0.9549
Proposed	0.9584	**0.9606**	**0.9563**

4.3 Limitations

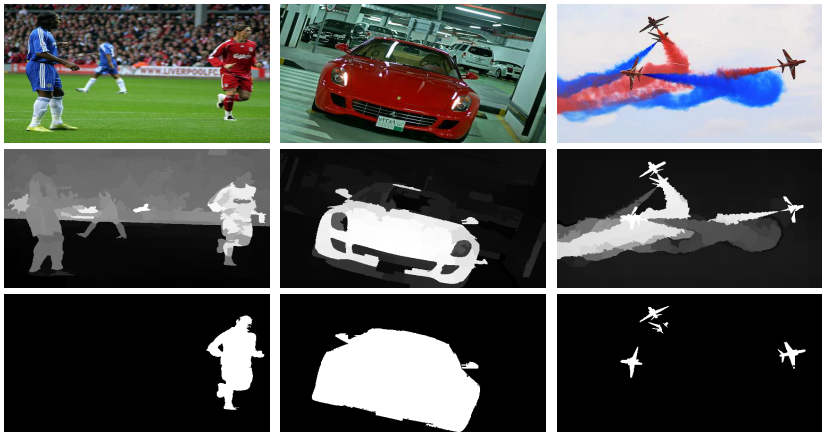


Figure 4.19: Limitations of the proposed approach (*from top to bottom*: input images, results of the proposed method and ground truth images, *from left to right*: *Liverpool*, *Ferrari* and *Air Show* sets).

There exist some failure cases of the proposed method, though extensive experiments verify that the proposed method is capable of detecting co-salient objects more precisely than the others on average. For example, some background and individual salient objects (blue players) are not fully suppressed on the *Liverpool* set (first column in Figure 4.19) because they commonly appear in some images, while some parts of salient object are mis-detected since the initial saliency detector based on the contrast prior fails to detect those regions *Ferrari* set (2nd column in Figure 4.19). As the proposed method uses a graph based on only color affinity, non-salient regions can be detected in case that it shares similar color with co-salient objects like red smoke in *Air Show* set (3rd column in Figure 4.19). These problems are expected to be alleviated in a certain extent by introducing relevant features along with color features.

Chapter 5

Specific Target: Skin Detection

5.1 Proposed Approach for Skin Detection

The proposed skin detection method is presented in this section, which consists of 5 steps: preprocessing, graph construction, multi seed extraction, propagation and region to pixel refinement. In the preprocessing step, the initial skin probability map (iSPM) for an input image is generated from the Bayesian classifier, and the input image is over-segmented. In the graph construction step, a multilayer graph is constructed with both image nodes and cluster nodes. In the third, two types of seeds are selected, which are from the iSPM and image boundaries respectively. With the prepared graph and the seeds, propagation is conducted which transfers the seeds of certain nodes to all the nodes to compute a skin probability map (SPM) for image nodes. Because the SPM represents skin probabilities for the over-segmented regions (nodes), a pixel-wise skin map is computed by measuring the similarities between the pixels and their corresponding image nodes. The details of the above steps are presented in the following subsections.

Figure 5.1: An illustration of each step for the proposed method (*from left to right*: input image, iSPM, extracted seeds on over-segmented regions, propagation result, and refinement result).

5.1.1 Preprocessing

Bayesian skin probability

In this subsection, the Bayesian skin color classifier [56] is briefly introduced since it is used to extract skin seeds and construct a graph in the proposed method. It is known that the Bayesian skin color classifier based on histograms detects skin pixels more accurately than other classical methods (e.g., Gaussian models, elliptical boundary models [55, 56]) as long as there are plenty of samples for training, and it is also very simple because it uses a look-up table when detecting the skin pixels.

Two histograms of $64 \times 64 \times 64$ bins is constructed as referring to [24], each of which respectively represents skin and non-skin histogram from a large training dataset. The histograms are normalized in order to make the sum of each histogram become one, which is given by $P(c|S)$ and $P(c|N)$ for skin and non-skin respectively. Assuming that the presence of skin and non-skin pixels is identical ($P(S) = P(N) = 0.5$), the Bayesian probability comes as:

$$\begin{aligned} P(S|c) &= \frac{P(c|S)P(S)}{P(c|S)P(S) + P(c|N)P(N)} \\ &= \frac{P(c|S)}{P(c|S) + P(c|N)}. \end{aligned} \tag{5.1}$$

A look-up table for skin probability is obtained from the above equation, and the

iSPM is generated. Note that other various skin detection methods can be employed in the proposed scheme.

Image segmentation

An image is divided into over-segmented regions based on the SLIC algorithm [64]. The over-segmented regions are denoted as $V^I = \{v_i^I\}_{i=1}^N$ where N is the number of over-segments, and each region is described by color and initial skin probability of the region as:

- $\mathbf{x}_i$: *Lab mean color vector* obtained by averaging color vectors of all pixels belonging to v_i^I,
- p_i: *mean skin probability* obtained by averaging iSPM of all pixels belonging to v_i^I.

The over segmentation step is employed for two reasons. First, the segmentation contributes to decreasing a lot of computational cost as the number of graph nodes becomes much smaller than that of pixel-wise case. Second, it can effectively represent the spatial relations between the nodes because a segment of pixels bears much spatial information, whereas the pixel-wise comparison sometimes fails to find the spatial relationship.

5.1.2 Graph Construction

A graph for the proposed method is constructed with a set of nodes V and a set of edges E connecting the pairs of nodes, which is denoted as $G = (V, E)$. Each edge in the graph is given strength (weight) that represents how much the paired nodes are similar to each other. In the proposed method, a two layer graph (image layer

and cluster layer) is constructed where each layer has different type of node from the other, denoted as $V = \{V^I, V^C\}$ where V^I is a set of the nodes in the image layer and V^C is a set of the nodes in the cluster layer. As the graph is composed of two layers, there are three types of edges connecting the nodes inside and inter layers, denoted as $E = \{E^I, E^C, E^{VC}\}$, which are the edges connecting the nodes inside the image layer, inside the cluster layer and between the image and cluster layer respectively. The image layer and the cluster layer are constructed separately at first, and then these two layers are connected together by E^{VC}.

For the image layer graph, the nodes are defined as the over-segments V^I which are connected by its neighbors with weight defined as:

$$w_{ij}^I = \exp\left(-\frac{|p_i - p_j|^2}{\sigma_p^2}\right), \tag{5.2}$$

where σ_p is a control parameter for intra node affinity. As in [73], the σ_p is designed to be adjusted according to the input image statistics, which is defined as:

$$\sigma_p = \sqrt{\frac{1}{|E^I|}\sum_{e_{ij}\in E^I}(p_i - p_j)^2}, \tag{5.3}$$

where $\left|E^I\right|$ is the number of intra edges. In the proposed method, an edges is linked with weight defined with a similarity of skin probability between nodes, while other methods including [11, 73] connect the edges according to color affinity. In the experiments section, it will be shown that the proposed approach provides better performance in reducing the false positive pixels because the color affinity tends to detect skin-like color nodes which are actually non-skin, such as reddish and yellow colors (see Figure 5.3 that will be explained in experiment section). From equation (5.2), an affinity matrix $\mathbf{W}_I$ is constructed whose (i, j)-th element is the weight

between i-th and j-th node, defined as:

$$(\mathbf{W}_I)_{i,j} = \begin{cases} w_{ij}^I, & \text{if } i \in Q_j, \\ 0, & \text{otherwise,} \end{cases} \tag{5.4}$$

where Q_j is a set of neighbor nodes for v_i^I, which implies that the nodes (over-segmented regions) in the image layer are sparsely connected to only its neighbors.

For the construction of cluster layer graph, the image nodes are firstly categorized into some groups by K-means clustering algorithm where the feature for the clustering process is the color vector of each node denoted as $\mathbf{x}_i$. This cluster is defined as a node in the cluster layer denoted as v_i^C, and define the set of these nodes as $V^C = \{v_i^C\}_{i=1}^K$, where K is the number of clusters. It is noted that the centroid of the i-th cluster is the average of all color vectors of the nodes belonging to the cluster, which is denoted as $\mathbf{c}_i$. The affinity for intra cluster graph is defined as the similarity between the centroids of clusters which is written as:

$$w_{ij}^C = \exp\left(-\frac{\|\mathbf{c}_i - \mathbf{c}_j\|}{\sigma_c}\right), \tag{5.5}$$

where σ_c is a control parameter for the similarity, which is a fixed value unlike the image graph case, because there are no sufficient connections between the clusters for computing statistics. From the above affinity, the affinity matrix $\mathbf{W}_C$ for the intra cluster is represented whose (i,j)-th is the affinity between i-th and j-th node, defined as:

$$(\mathbf{W}_C)_{i,j} = \begin{cases} w_{ij}^C, & \text{if } i \in k\text{-NN}(j) \text{ or } j \in k\text{-NN}(i), \\ 0, & \text{otherwise,} \end{cases} \tag{5.6}$$

where the k-NN(j) and k-NN(i) are the k-nearest neighbors for the nodes v_j^C and v_i^C, which means that it is also constructed by linking the nodes sparsely to their neighbors in the *Lab* color space.

Now, two graphs are integrated by connecting the image layer graph and the cluster layer graph in the form of a multilayer graph. That is, the nodes in the two graphs are connected by the inter-edges. The similarity between an image node and a cluster node is similarly defined as equation (5.5), which is written as:

$$w_{ij}^{IC} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{c}_j\|}{\sigma_c}\right), \tag{5.7}$$

where σ_c is a color control parameter for the inter cluster similarity. In the proposed method, each image node is connected to only one cluster node and thus the (i, j)-th element of the affinity matrix $\mathbf{W}_{IC}$ is represented as:

$$(\mathbf{W}_{IC})_{i,j} = \begin{cases} w_{ij}^{IC}, & \text{if } i \in C_j, \\ 0, & \text{otherwise}, \end{cases} \tag{5.8}$$

where C_j is a set of image layer nodes included in the cluster, v_j^C.

Finally, the overall graph representation using an affinity matrix can be expressed as a block-wise matrix form:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_I & \mathbf{W}_{IC} \\ \mathbf{W}_{CI} & \mathbf{W}_C \end{bmatrix}, \tag{5.9}$$

where $\mathbf{W}_{CI}$ is the transpose of $\mathbf{W}_{IC}$.

5.1.3 Seed Extraction

For the semi-supervised learning [12], it is necessary to select the seeds whose labels are known from a simple preceding classification process or from some assumptions and/or prior knowledge. As stated in the introduction, one of the contributions of the proposed method is to prepare two kinds of seeds, i.e., skin seeds and non-skin seeds whereas the conventional methods using seed propagation use only the former

one. The proposed method selects skin seeds as a set of nodes with high iSPM while non-skin seeds as a set of nodes adjacent to image boundaries as well as the ones with low iSPM. It is based on two prior knowledge, one of which is that the initial simple skin detection algorithm (Bayesian detector with color statistics in the proposed method) works well in a certain extent, and the other is that humans are generally foreground (main) objects in images and are located around the center of image which are general assumptions in the saliency research.

For the skin seeds, K-means clustering is conducted to make three groups, based on the features as a set of skin probability, $\{p_i\}_{i=1}^{N}$ of all nodes. Then, there are three clusters whose centroids represent the skin probability of clusters. Nodes belonging to the most skin-probable cluster are selected as skin seeds which should be not located at image boundaries, while nodes belonging to the least skin-probable cluster are extracted as non-skin seeds which should be located at image boundaries. The skin and the non-skin seeds are notated as $\hat{\mathbf{y}}_S$ and $\hat{\mathbf{y}}_B$ respectively.

5.1.4 Seed Propagation

Skin detection based on multi-seed propagation is explained in detail. The proposed method exploits a semi-supervised learning algorithm [12] which is applied to many saliency detection methods [11, 73]. Because these methods aim at classifying most unknown samples from a few known samples, the framework is applied to the proposed skin detection task. Referring to [12], a matrix for the semi-supervised learning is defined as:

$$\mathbf{W}_L = (1-\alpha)\left(\mathbf{D} - \alpha\mathbf{W}\right)^{-1}, \tag{5.10}$$

where α is a regularization control parameter, and $\mathbf{D}$ is the degree matrix of $\mathbf{W}$ where each diagonal entry is a row sum of $\mathbf{W}$. As a result, $\mathbf{W}_L$ becomes a fully

connected graph even though the original graph ($\mathbf{W}$) is connected sparsely at first.

Propagation is conducted based on the graph ($\mathbf{W}_L$) along with both skin seeds ($\mathbf{y}_S$) and background seeds ($\mathbf{y}_B$). It is necessary that the seed vectors are filled by zero vectors to adjust the size of the vectors, because the graph is composed of image nodes and cluster nodes, while there are only image nodes for the seed vectors, which is given by $\hat{\mathbf{y}}_S$ and $\hat{\mathbf{y}}_B$ respectively. Then, scores of being a skin node and being a non-skin are computed by propagating both seeds to all the nodes which is simply defined as:

$$\hat{\mathbf{f}}_S = \mathbf{W_L}\hat{\mathbf{y}}_S, \tag{5.11}$$

$$\hat{\mathbf{f}}_B = \mathbf{W_L}\hat{\mathbf{y}}_B, \tag{5.12}$$

which are interpreted as the overall affinities to the skin and the background seeds respectively. A skin probability map is computed from these scores as in chapter 3, which is written as:

$$\hat{\mathbf{p}}_S = (\hat{\mathbf{f}}_S - \hat{\mathbf{f}}_B)./(\hat{\mathbf{f}}_S + \hat{\mathbf{f}}_B), \tag{5.13}$$

where the ./ is a element-wise division operator. Then, skin probability for the region, v_i^I, which is the i-th element of $\hat{\mathbf{p}}_S$, is notated as $P(v_i^I)$.

5.1.5 Pixel-wise Refinement

From the propagation, the SPM for the over-segments is available, not for the pixels itself. Since the segmentation algorithm is not perfect and/or due to the nature of over-segmentation method, the segmented regions sometimes contain the pixels which are much different from the others in the same segment, for example the eyes in a segment, nostrils, mouths and so on. In this case, the SPM for the over-segments is lack of representing skin regions and non-skin regions. To alleviate this problem, a

region (over-segment) to pixel refinement process is conducted with color similarity measure between them which plays a role in suppressing the small non-skin pixels included in skin regions. The measure is defined as below:

$$P(\mathbf{x}_i^p|v_j^I) = \exp\left(-\frac{\|\mathbf{x}_i^p - \mathbf{x}_j\|^2}{\sigma_r^2}\right), \tag{5.14}$$

where $\mathbf{x}_i^p$ is a color vector for the i-th pixel and σ_r is a control parameter for the refinement. Pixel-wise skin probability is computed from the above equation and $P(v_j^I)$ as

$$P(\mathbf{x}_i^p) = P(\mathbf{x}_i^p|v_j^I) \cdot P(v_j^I). \tag{5.15}$$

5.2 Experiments

In this section, the results of extensive experiments are presented to validate the effectiveness of proposed method and to compare the performance with existing methods. First, the parameters used for the proposed method are determined for the entire experiments. Then, the proposed method is analyzed by varying some structures of the algorithm, e.g., affinity for the intra image nodes, the effect of the cluster nodes and the refinement process. Finally, the proposed method is compared with various algorithms to show the competitiveness of the proposed method.

5.2.1 Experiment Setup

There are some parameters for the proposed method which are decided empirically. The number of over-segmented regions is set to 250, and the K used in K-means clustering for cluster nodes is 50, while the number of nearest neighbors used for connecting the cluster layer is set to 5. The control parameters for the cluster affinity (σ_c) and refinement (σ_r^2) are determined to 0.05 and 0.1 respectively. As

in [11, 12, 73], the regularization control parameter, α, is set to 0.99. The Bayesian skin classifier was trained from $2,000$ images among ECU dataset [74] which consists of $4,000$ images.

The performance of the proposed method is evaluated on three datasets: ECU [74], HGR [24] and Pratheepan [75]. The ECU is a set of images with some parts of human including faces or whole bodies taken from various illumination conditions, which is the largest data set for the purpose of skin detection research. It is noted that the experiments are conducted on a subset of ECU dataset ($2,000$ images) because some of the set are exploited in the training process for the Bayesian classifier. The HGR dataset is created for hand gesture recognition consisting of 899 images with several hand gestures of 12 different individuals, which is regarded comparably easier than other dataset. The Pratheepan is 32 facial images, each of which includes only one person with the face. It is suggested for skin detection based on facial region detectors by *Yogarajah et al.* [75]. Precision-recall curve (PR curve) and receiver operating characteristic curve (ROC curve) are used for overall performance evaluation. In addition, the performance of binarized skin maps is evaluated with four statistics: *Accuracy*, *Precision*, *Recall* and *F-measure* in a fixed threshold which maximizes the F-measure of each method. It is noted that ROC is computed from false positive rate and false negative rate, and β^2 is set to 1 for F-measure in equation (3.25) because these are widely used in skin detection research.

5.2.2 Analysis of the Proposed Approach

Some experiments are carried on to investigate the effectiveness of internal set-up in the proposed method. First, two cases for intra image affinity are compared, one is the case of connecting the edges with color similarity and the other connecting

the edges with SPM. It is shown from Figure 5.2 that the SPM case shows better performance in the overall ranges on both PR and ROC curves than the color case. Unlike the saliency task [11, 73] which relies on bottom-up features from input images, it is possible to exploit the top-down features from the learning in the case of skin detection. As a result, reddish and other skin-like colors are not detected in the SPM case as shown in Figure 5.3 (6th column compared to 3rd column).

An experiment for testing the effectiveness of the cluster layer has also been conducted by comparing the cases with/without the cluster layer. It is quantitatively demonstrated in Figure 5.2 that introducing the cluster nodes contributes to the increase of detection performance. Some skin regions are suppressed because these are so spatially isolated from the large skin regions, while some background regions are not suppressed as shown in the 4th row of Figure 5.3. However, the problem is alleviated by connecting those regions via cluster nodes which possibly link similar color regions together even though it is not spatially adjacent (6th column of Figure 5.3).

Lastly, the effectiveness of the pixel-wise refinement process is investigated. Since the proposed method starts by segmenting the image, it cannot well reflect the characteristics of individual pixels. Hence, small objects (e.g., eyes in 5th column of Figure 5.3) included in the segments are also detected as skin pixels. The pixel-wise refinement process improves the detection performance by considering region-pixel coherence, which is also presented in Figure 5.2.

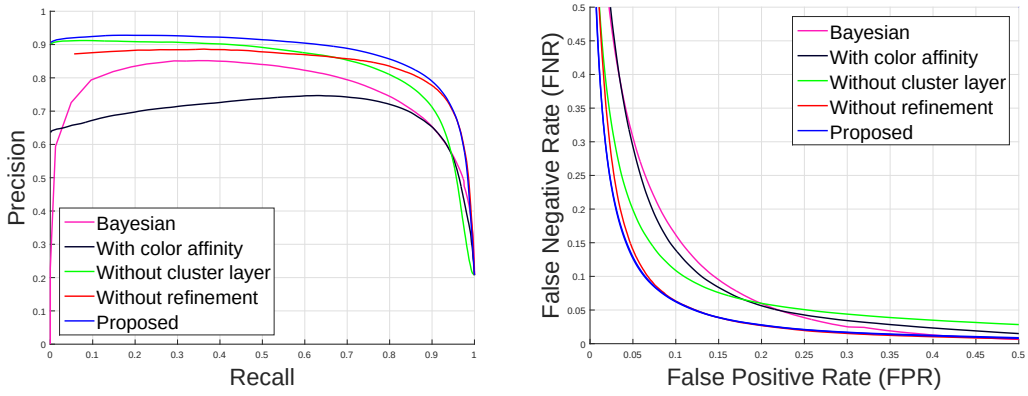


Figure 5.2: Quantitative comparison of the proposed method with variations: PR curves (left) and ROC curves (right).

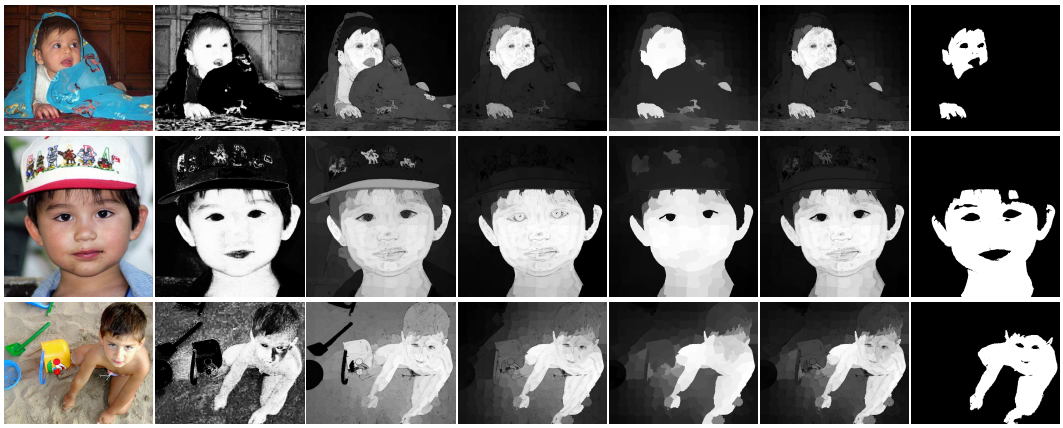


Figure 5.3: Visual comparison of the proposed method with variations (*from left to right*: input images, results of the Bayesian classifier, results with the color affinity, results without the cluster layer, results without the pixel-wise refinement, results of the proposed method and ground truth images).

5.2.3 Comparison with Other Algorithms

In this subsection, the proposed method is compared with other methods: Bayesian [56], FPSD [23], DSPF [24], FSD [62] and LASD [63]. The Bayesian classifier is compared in order to provide a baseline for the proposed algorithm, because it is used for the skin seed extraction and graph construction steps of the proposed method. FPSD obtains seeds from a conventional skin detector, and propagate it to the whole image based on the distance transform. DSPF extracts more reliable features from texture information for the propagation of seeds over the graph. Two skin detectors (FSD, LASD) based on on-line training have been also compared, which obtain the samples from the facial regions via some face-feature detectors (e.g., eye detector) and exploit these samples to make adaptive skin models. FSD mixes the dynamic threshold model and Gaussian model together in skin detection while LASD models luminance adaptive color space using least square error optimization.

The overall performance of the compared methods are plotted on PR curves and ROC curves in Figure 5.4, 5.5 and 5.6 respectively. For the ECU dataset, the proposed method locates the PR curves over those of other methods and also locates the ROC curves under the others. Also, the curves for the HGR shows that the proposed method is comparable or better in most ranges. It shows that the proposed propagation approach has an advantage of accurately detecting skin pixels than the propagation of FPSD method. Furthermore, the proposed method yields better results than DSPF which is significantly improved by introducing texture features along with the propagation of FPSD. The reason is that two types of seeds are exploited, especially one of which is extracted from image boundaries. The proposed method is expected to be considerably improved when adopting more

relevant seed extraction method as DSPF.

In addition, the quality of binary skin maps has been investigated by applying a threshold. The threshold for each method is selected in accordance with maximizing the F-measure which represents a classification performance well-balanced for precision and recall. It is shown from Table 5.1 that the proposed method also shows better performance in most measures on the ECU and the HGR datasets. The proposed method is also compared with FSD and LASD along with other methods on the Pratheepan dataset as shown in Table. 5.2 because the set consists of facial images only, and those methods are capable of using the detector for facial features. Though FSD record the highest score in precision because their objective is mainly to reduce the false positives, the proposed method outperforms other methods in terms of accuracy, recall and F-measure without adopting any other detector.

Visual comparison is presented in Figure 5.7. The proposed method is shown to be robust in suppressing the background (non-skin) regions compared to other propagation methods due to the background seeds, while detecting more skin pixels than detector based methods.

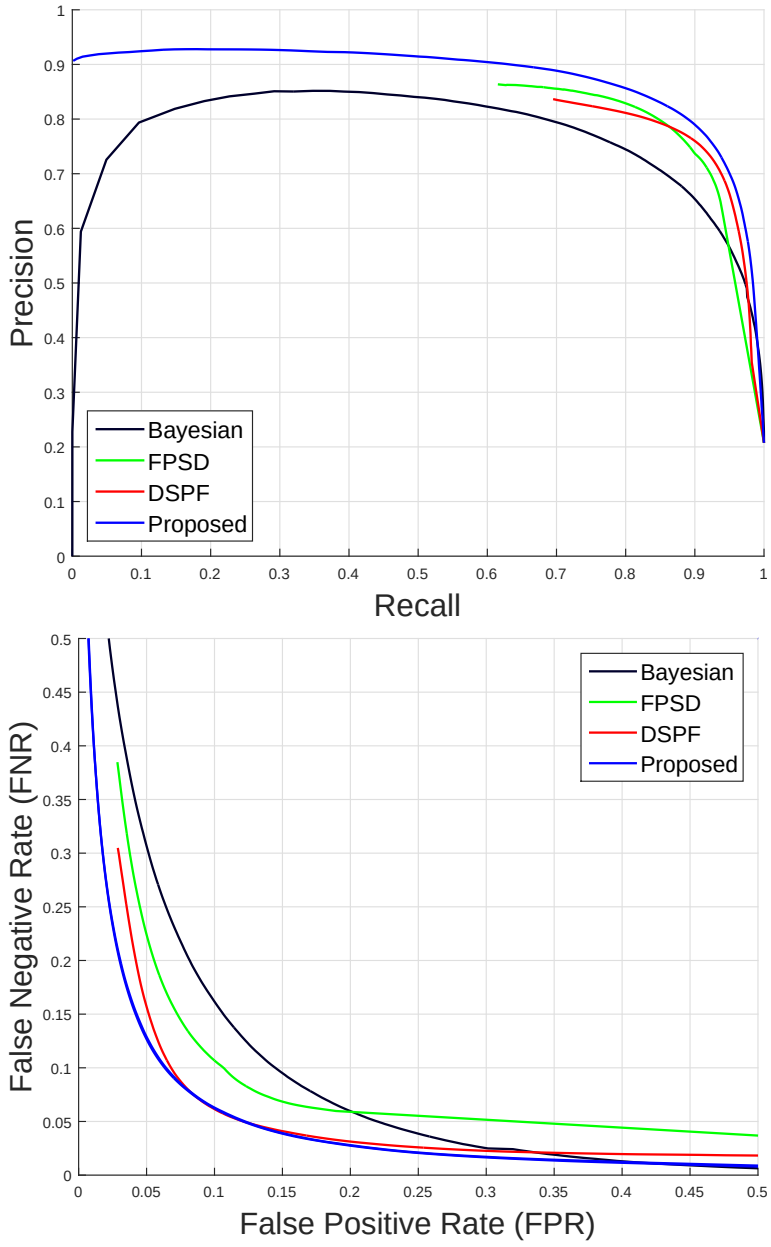


Figure 5.4: Quantitative comparison on ECU dataset: PR curves (top) and ROC curves (bottom).

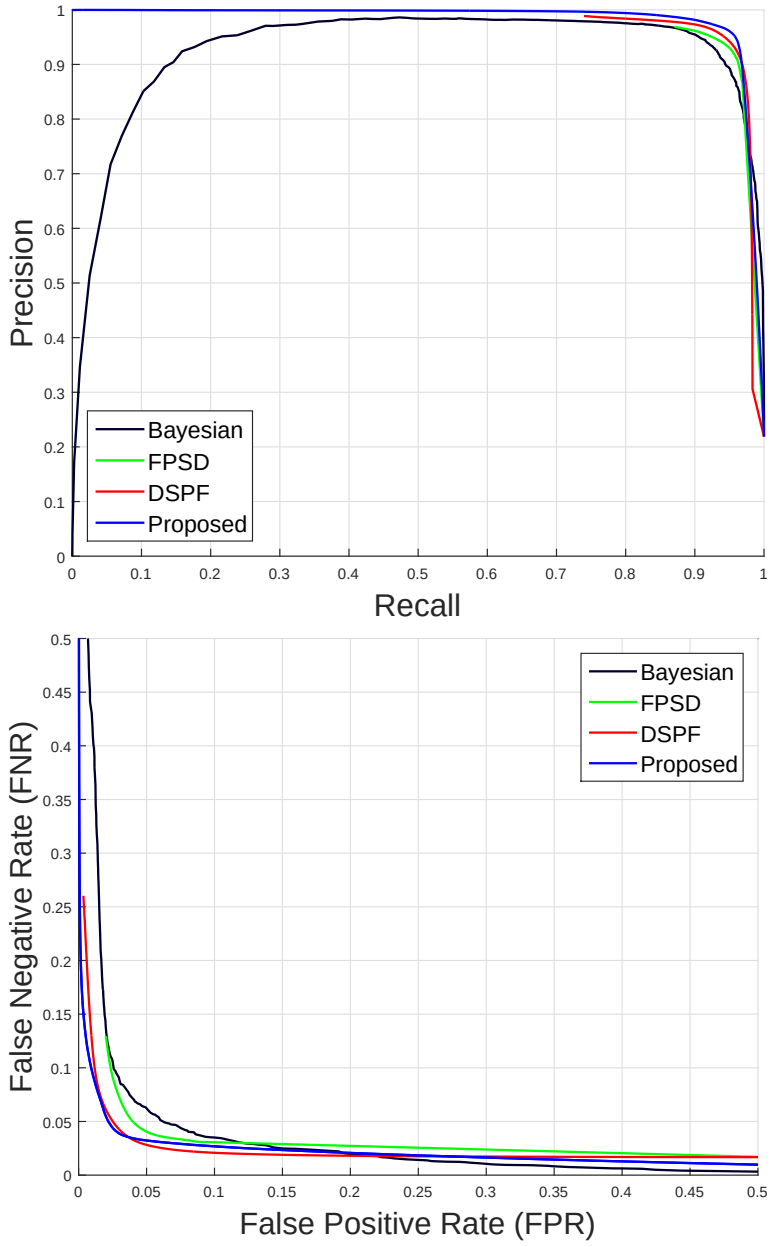


Figure 5.5: Quantitative comparison on HGR dataset: PR curves (top) and ROC curves (bottom).

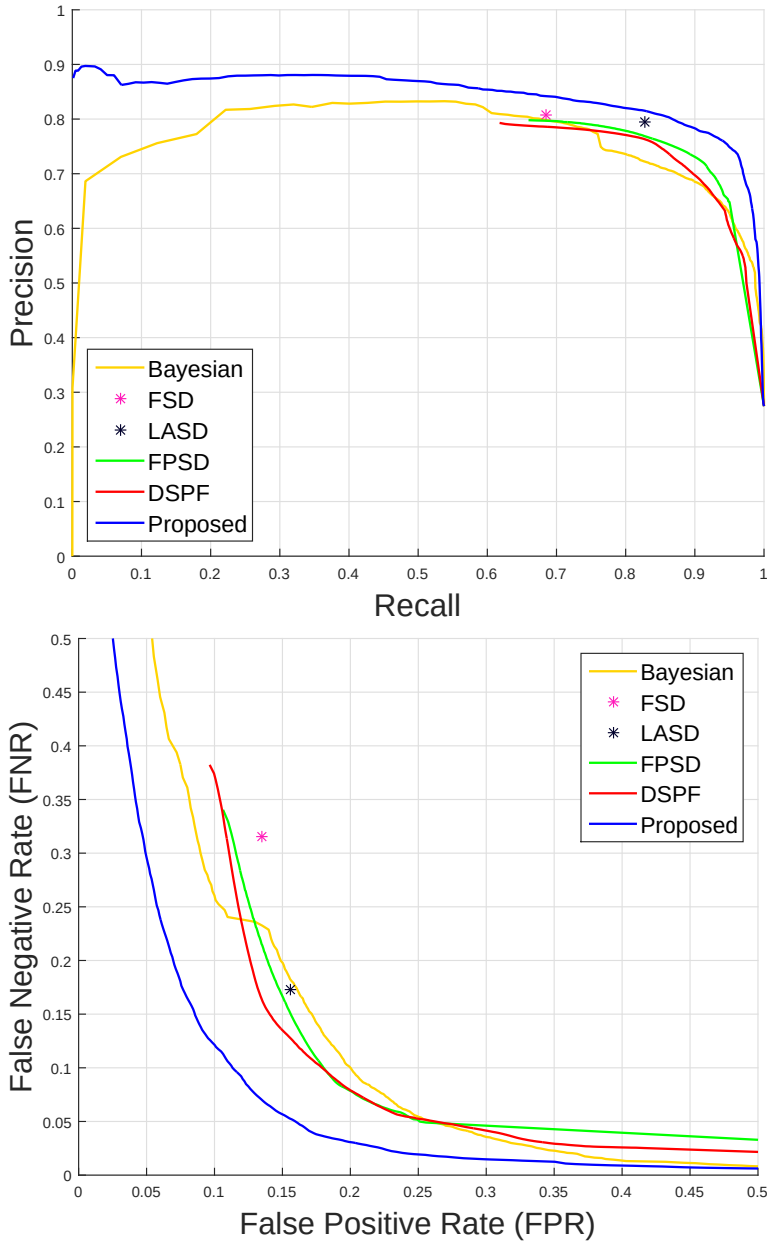


Figure 5.6: Quantitative comparison on Pratheepan dataset: PR curves (top) and ROC curves (bottom).

Table 5.1: Accuracy, precision, recall and F-measure at peak F-measure on ECU and HGR datasets.

Methods	ECU dataset [74]				HGR dataset [23, 24]			
	Accuracy	*Precision*	*Recall*	*F-measure*	*Accuracy*	*Precision*	*Recall*	*F-measure*
Bayesian [56]	0.8910	0.7292	0.8220	0.7728	0.9598	0.9447	0.9152	0.9297
FPSD [23]	0.9106	0.7948	0.8534	0.8231	0.9610	0.9348	0.9458	0.9403
DSPF [24]	0.9190	0.7713	**0.8864**	0.8249	0.9701	0.9494	0.9437	0.9465
Proposed	**0.9306**	**0.8085**	0.8805	**0.8430**	**0.9735**	**0.9593**	**0.9517**	**0.9555**

Table 5.2: Accuracy, precision, recall and F-measure at peak F-measure on Pratheepan dataset.

	Accuracy	*Precision*	*Recall*	*F-measure*
Bayesian [56]	0.8237	0.6881	0.8972	0.7788
FSD [62]	0.8255	**0.8077**	0.6851	0.7414
LASD [63]	0.8361	0.7954	0.8275	0.8111
FPSD [23]	0.8419	0.7387	0.8991	0.8070
DSPF [24]	0.8521	0.7543	0.8436	0.7964
Proposed	**0.8782**	0.7659	**0.9328**	**0.8412**

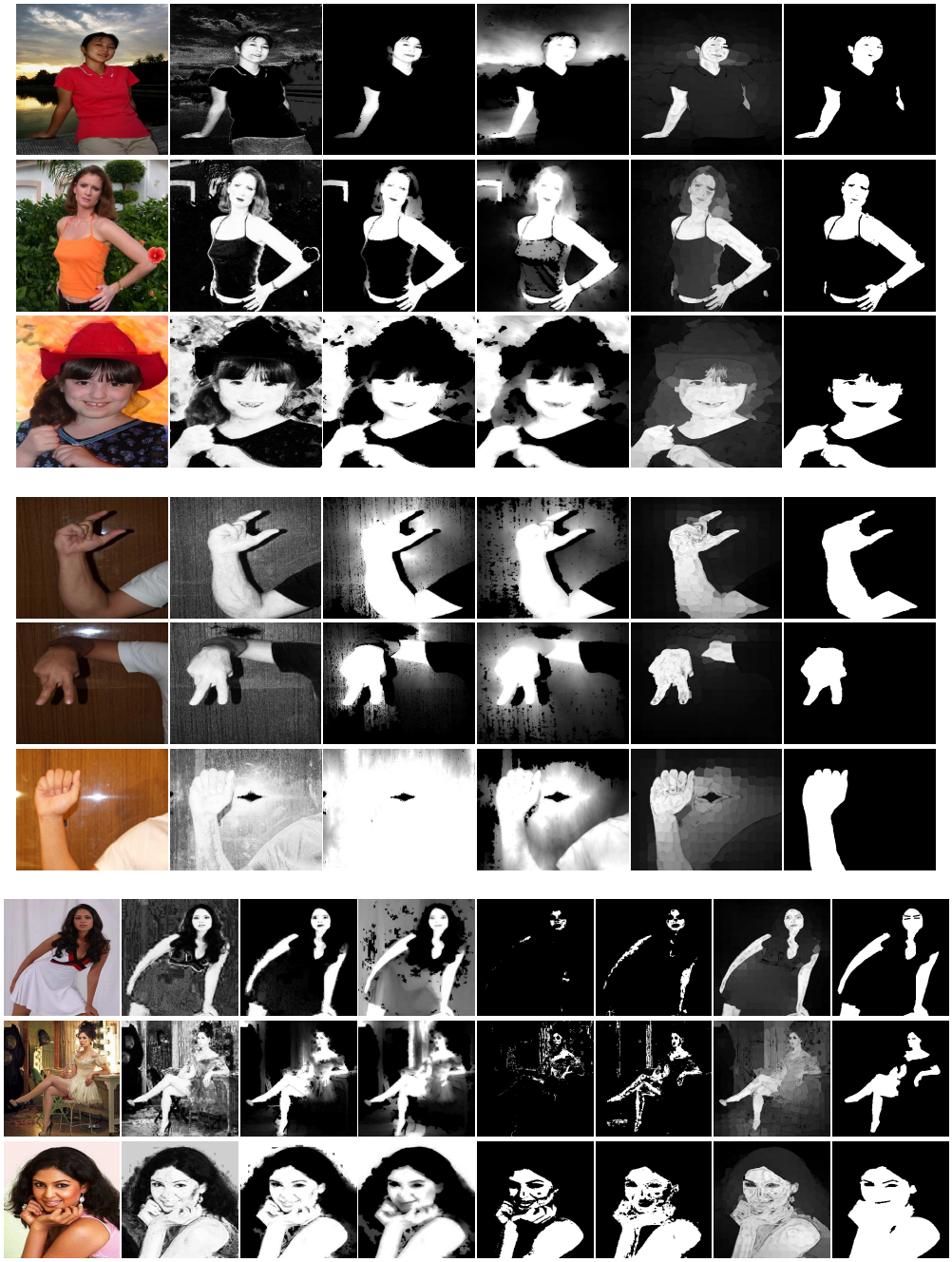


Figure 5.7: Visual comparison on three datasets (*from left to right*: input, Bayesian, FPSD, DSPF, FSD, LASD, proposed, and ground truth images).

Chapter 6

Conclusion

In this dissertation, saliency detection methods have been proposed, by means of propagation on multilayer graphs learned by a semi-supervised learning scheme with multi seeds of different types. The multi seeds are introduced to effectively suppress background regions, which significantly reduces false positive errors. Meanwhile, multilayer graphs are constructed to maintain the consistency of saliency maps in the way that it uniformly highlights salient objects and background regions. The proposed approach has been adaptively used for various subjects, i.e., saliency detection, co-saliency detection, and skin detection in this dissertation, each of which is concluded in the following subsections.

6.1 Saliency Detection for Single Images

A saliency detection algorithm has been proposed based on the graph representation of over-segmented image and seed propagation over the graph. The contribution of the proposed method is that both saliency seeds and boundary background seeds

are exploited for the propagation, and the saliency is computed by comparing the propagation scores from both seeds. In addition, two different multilayer graph models have been introduced, one of which includes the higher layer from a coarsely segmented image for spatial regularization, and another model builds the graph along with the color cluster layer that maintains the coherence of saliency among the regions sharing similar colors. Experiments have shown that the proposed approach yields better results on various objective measures (e.g., PR curve, ROC curve, F-measure and AUC), and also makes visually plausible results on various datasets.

6.2 Co-saliency Detection for Multiple Images

A co-saliency detection algorithm has also been proposed, adopting a basically similar framework for the proposed saliency detection approach. Meanwhile, coherence cue is additionally applied for the co-saliency seed selection in order to extract regions that are commonly salient in all images. The proposed method builds a multilayer graph which consists of intra image layer for each image and cluster layer. The intra image layers are constructed in the same way of saliency detection. However, the cluster layer plays a role in indirectly connecting the image nodes of different layers as well as the nodes of same layers, because it is hard to directly connect the nodes between different image layers which have co-salient objects and background of different scales and location. Furthermore, the cluster layer is to maintain the consistency of co-salient objects among images in the way that it connects nodes sharing high similarity together regardless of intra or inter nodes. Experiments have shown that the proposed method outperforms in the quantitative measures used for the evaluation of single image saliency, and provides visually more superior images

in the respect to homogeneity of the objects on all datasets.

6.3 Skin Detection

A skin detection method has been proposed, based on the multi-layer graph representation of an image and multi-seed propagation over the graph. The input image is represented by a two-layer graph, where the nodes in the first layer are the over-segmented regions and the nodes in the second are the clusters of these segments. The graph is designed in such a way that the connectivity of nodes represents the color similarity and also the spatial relationship. In summary, the contribution of the proposed method is to design a multi-layer graph that can deal with color similarity as well as spatial relations for the detection of skin pixels, and also to devise a multi-seed propagation method whereas the conventional graph based method considered only one kind of seed. Specifically, the conventional methods considered only skin seed while our method prepares skin and non-skin (background) seeds where the latter plays an important role in suppressing the background regions and consequently reducing false positive errors. Experiments have shown that the proposed method performs better than others even though the proposed method does not use texture based seed extraction or facial detectors.

Bibliography

[1] A. Borji and L. Itti, "Scene classification with a sparse set of salient regions," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2011, pp. 1902–1908.

[2] Z. Liang, H. Fu, Z. Chi, and D. Feng, "Image pre-classification based on saliency map for image retrieval," in *Proc. IEEE Int. Conf. Information, Communications and Signal Processing*, 2009, pp. 1–5.

[3] V. Setlur, T. Lechner, M. Nienhaus, and B. Gooch, "Retargeting images and video for preserving information saliency," *IEEE Comput. Graph. Appl.*, vol. 27, no. 5, pp. 80–88, Sep. 2007.

[4] Y. Fang, Z. Chen, W. Lin, S. Member, and C.-W. Lin, "Saliency detection in the compressed domain for adaptive image retargeting," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3888–3901, Sep. 2012.

[5] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 185–198, Jan. 2010.

[6] Y.-F. Ma, X.-S. Hua, L. Lu, and H.-J. Zhang, "A generic framework of user attention model and its application in video summarization," *IEEE Trans. Multimedia*, vol. 7, no. 5, p. 907, Oct. 2005.

[7] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs for salient object detection in images," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3232–3242, Dec. 2010.

[8] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing markov chain," in *Proc. IEEE Int. Conf. Computer Vision*, 2013, pp. 1665–1672.

[9] T. H. Kim, K. M. Lee, and S. U. Lee, "Learning full pairwise affinities for spectral segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1690–1703, Jul. 2013.

[10] S. Lu, V. Mahadevan, and N. Vasconcelos, "Learning optimal seeds for diffusion-based salient object detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2014, pp. 2790–2797.

[11] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2013, pp. 3166–3173.

[12] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Proc. Advances in Neural Information Processing Systems*, 2004, pp. 321–328.

[13] L. Zhou, R. Hu, Y. Qian, and H.-H. Chen, "Energy-spectrum efficiency tradeoff for video streaming over mobile ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 981–991, May 2013.

[14] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2011, pp. 409–416.

[15] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.

[16] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. European Conf. Computer Vision*, 2012, pp. 29–42.

[17] G. Hu, Y. Zhao, Y. Yuan, and D. Gu, "Local stereo matching with canny segmentation and reliable seed propagation," in *Proc. Int. Conf. Computer Vision in Remote Sensing*, 2012, pp. 177–182.

[18] J.-S. Kim, J.-Y. Sim, and C.-S. Kim, "Multiscale saliency detection using random walk with restart," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 2, pp. 198–210, Feb. 2014.

[19] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2013, pp. 1155–1162.

[20] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2014, pp. 2814–2821.

[21] M.-M. Cheng, N. J. Mitra, X. Huang, and S.-M. Hu, "Salientshape: Group saliency in image collections," *Vis. Comput.*, vol. 30, no. 4, pp. 443–453, Apr. 2014.

[22] J. Ruiz-del Solar and R. Verschae, "Skin detection using neighborhood information," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2004, pp. 463–468.

[23] M. Kawulok, "Fast propagation-based skin regions segmentation in color images," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2013, pp. 1–7.

[24] M. Kawulok, J. Kawulok, and J. Nalepa, "Spatial-based skin detection using discriminative skin-presence features," *Pattern Recognition Letters*, vol. 41, pp. 3–13, 2014.

[25] S. Ramanathan, H. Katti, N. Sebe, M. Kankanhalli, and T.-S. Chua, "An eye fixation database for saliency detection in images," in *Proc. European Conf. Computer Vision*, 2010, pp. 30–43.

[26] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Proc. IEEE Int. Conf. Computer Vision*, 2009, pp. 2106–2113.

[27] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *Proc. Advances in Neural Information Processing Systems*, 2005, pp. 155–162.

[28] L. Itti and P. F. Baldi, "Bayesian surprise attracts human attention," in *Proc. Advances in Neural Information Processing Systems*, 2005, pp. 547–554.

[29] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[30] X. Hou and L. Zhang, "Saliency detection : A spectral residual approach," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[31] C. Yang, L. Zhang, and H. Lu, "Graph-regularized saliency detection with convex-hull-based center prior," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 637–640, Jul. 2013.

[32] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[33] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2009, pp. 1597–1604.

[34] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *Proc. IEEE Int. Conf. Computer Vision*, 2013, pp. 1529–1536.

[35] D. E. Jacobs, D. B. Goldman, and E. Shechtman, "Cosaliency: Where people look when comparing images," in *Proc. Symp. User Interface Software and Technology*, 2010, pp. 219–228.

[36] H. Li and K. N. Ngan, "A co-saliency model of image pairs," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3365–3375, Dec. 2011.

[37] Z. Tan, L. Wan, W. Feng, and C.-M. Pun, "Image co-saliency detection by propagating superpixel affinities," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 2013, pp. 2114–2118.

[38] X. Cao, Z. Tao, B. Zhang, H. Fu, and X. Li, "Saliency map fusion based on rank-one constraint," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2013, pp. 1–6.

[39] K.-Y. Chang, T.-L. Liu, and S.-H. Lai, "From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2011, pp. 2129–2136.

[40] X. Cao, Z. Tao, B. Zhang, H. Fu, and W. Feng, "Self-adaptively weighted co-saliency detection via rank constraint," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4175–4186, Sep. 2014.

[41] H. Fu, X. Cao, and Z. Tu, "Cluster-based co-saliency detection," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3766–3778, Oct. 2013.

[42] H. Li, F. Meng, and K. N. Ngan, "Co-salient object detection from multiple images," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1896–1909, Dec. 2013.

[43] Z. Liu, W. Zou, L. Li, L. Shen, and O. Le Meur, "Co-saliency detection based on hierarchical segmentation," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 88–92, Jan. 2014.

[44] L. Li, Z. Liu, W. Zou, X. Zhang, and O. Le Meur, "Co-saliency detection based on region-level fusion and pixel-level refinement," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2014, pp. 1–6.

[45] R. Huang, W. Feng, and J. Sun, "Saliency and co-saliency detection by low-rank multiscale fusion," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2015, pp. 1–6.

[46] Y. Li, K. Fu, Z. Liu, and J. Yang, "Efficient saliency-model-guided visual co-saliency detection," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 588–592, May 2015.

[47] B. D. Zarit, B. J. Super, and F. K. Quek, "Comparison of five color models in skin pixel classification," in *Int. Workshop Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, 1999, pp. 58–63.

[48] A. Cheddad, J. Condell, K. Curran, and P. M. Kevitt, "A new colour space for skin tone detection," in *Proc. IEEE Int. Conf. Image Processing*, 2009, pp. 497–500.

[49] A. Albiol, L. Torres, and E. J. Delp, "Optimum color spaces for skin detection." in *Proc. IEEE Int. Conf. Image Processing*, 2001, pp. 122–124.

[50] M. C. Shin, K. I. Chang, and L. V. Tsap, "Does colorspace transformation make any difference on skin detection?" in *IEEE Workshop Applications of Computer Vision*, 2002, pp. 275–279.

[51] G. Gomez and E. Morales, "Automatic feature construction and a simple rule induction algorithm for skin detection," in *ICML Workshop Machine Learning in Computer Vision*, 2002, pp. 31–38.

[52] M. Abdullah-Al-Wadud, M. Shoyaib, and O. Chae, "A skin detection approach based on color distance map," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 1, pp. 1–10, 2009.

[53] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 696–706, 2002.

[54] M.-H. Yang and N. Ahuja, "Gaussian mixture model for human skin color and its applications in image and video databases," in *SPIE: Storage and Retrieval for Image and Video Databases VII*, vol. 3656, 1998, pp. 458–466.

[55] J. Y. Lee and S. I. Yoo, "An elliptical boundary model for skin color detection," in *Int. Conf. Imaging Science, Systems, and Technology*, 2002.

[56] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *Int. J. Comput. Vision*, vol. 46, no. 1, pp. 81–96, 2002.

[57] R. Khan, A. Hanbury, and J. Stoettinger, "Skin detection: A random forest approach," in *Proc. IEEE Int. Conf. Image Processing*, 2010, pp. 4613–4616.

[58] Q. Zhu, K.-T. Cheng, C.-T. Wu, and Y.-L. Wu, "Adaptive learning of an accurate skin-color model," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2004, pp. 37–42.

[59] J. Geng, Z. Miao, and C. Zhong, "Skin detection with illumination adaptation in single image," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2011, pp. 1–6.

[60] C.-C. Hsieh, D.-H. Liou, and M.-K. Jiang, "Fast enhanced face-based adaptive skin color model," in *Int. Conf. Image Processing and Pattern Recognition in Industrial Engineering*, 2010, pp. 782 027–782 027.

[61] P. Yogarajah, J. Condell, K. Curran, P. McKevitt, and A. Cheddad, "A dynamic threshold approach for skin tone detection in colour images," *Int. J. Biometrics*, vol. 4, no. 1, pp. 38–55, 2011.

[62] W. R. Tan, C. S. Chan, P. Yogarajah, and J. Condell, "A fusion approach for efficient human skin detection," *IEEE Trans. Ind. Informat.*, vol. 8, no. 1, pp. 138–147, 2012.

[63] I. Hwang, S. H. Lee, B. Min, and N. I. Cho, "Luminance adapted skin color modeling for the robust detection of skin areas," in *Proc. IEEE Int. Conf. Image Processing*, 2013, pp. 2622–2625.

[64] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[65] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 353–367, Feb. 2011.

[66] A. Borji, D. Sihite, and L. Itti, "Salient object detection: A benchmark," in *Proc. European Conf. Computer Vision*, 2012, pp. 414–429.

[67] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proc. Int. Conf. Machine Learning*, 2006, pp. 233–240.

[68] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in *Proc. Int. Conf. Computer Vision Systems*, 2008, vol. 5008, pp. 66–75.

[69] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Advances in Neural Information Processing Systems*, 2007, pp. 545–552.

[70] D. S. Hochbaum and V. Singh, "An efficient algorithm for co-segmentation," in *Proc. IEEE Int. Conf. Computer Vision*, 2009, pp. 269–276.

[71] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "icoseg: Interactive co-segmentation with intelligent scribble guidance," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2010, pp. 3169–3176.

[72] ——, "Interactively co-segmentating topically related images with intelligent scribble guidance," *Int. J. Comput. Vision*, vol. 93, no. 3, pp. 273–292, 2011.

[73] I. Hwang, S. H. Lee, J. S. Park, and N. I. Cho, "Saliency detection based on seed propagation in a multilayer graph," *Multimedia Tools and Applications*, pp. 1–19, 2016.

[74] S. L. Phung, A. Bouzerdoum, and D. Chai, "Skin segmentation using color pixel classification: analysis and comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 1, pp. 148–154, 2005.

[75] P. Yogarajah, J. Condell, K. Curran, A. Cheddad, and P. McKevitt, "A dynamic threshold approach for skin segmentation in color images," in *Proc. IEEE Int. Conf. Image Processing*, 2010, pp. 2225–2228.

초록

사람은 시야 내에서의 영역 또는 객체에 대한 우선 순위를 부여함으로써 복잡한 환경에서도 시각적 정보를 아주 빠르게 이해하는 능력이 있다. 이미지 중요 객체 검출은 이미지 내에서의 중요한 객체를 검출해주는 것으로, 이는 컴퓨터 비전 및 영상 처리 분야의 다양한 연구에서 전처리(preprocessing)로 활용되어 매우 중요한 연구 주제이다. 최근에는 중요 객체나 영역을 대략적인 위치로 검출하기 보다는, 객체 단위로서 균일하게 검출하기 위하여 다양한 노력들이 이루어지고 있다. 본 논문에서는 다층 그래프 상에 씨드(seed)를 전파함으로써 중요 객체를 보다 정확하게 검출하는 접근법을 제안하고, 이를 단일 이미지 중요 객체 검출(saliency detection), 복수 이미지에서의 공통된 중요 객체 검출(co-saliency detection), 피부 검출(skin detection)에 각각 적용한다.

우선은 단일 이미지에서 씨드 전파를 활용한 중요 객체 검출 방법을 제안한다. 기존의 연구와는 달리, 제안하는 방법에서는 상반된 두 가지의 씨드(중요 객체를 나타내는 씨드와 배경을 나타내는 씨드)를 이용한다. 두 종류의 씨드는 준 지도 학습(semi-supervised learning) 기법을 활용하여 각각 따로 그래프를 통하여 전파되고, 전파된 결과를 이용하여 중요 객체를 검출한다. 또한 이러한 접근법을 다층 그래프로 확장시켜 객체를 보다 균일하게 검출하도록 한다. 본 논문에서는 두 가지의 다층 그래프를 활용한 검출 방법을 제안하는데, 하나는 공간적인 관계를 고려하여 중요 객체를 검출하는 방법이고, 다른 하나는 특징 공간 (색상 공간)에서의 관계를 고려하여 중요 객체를 검출하는 방법이다. 두 가지 방법은 단일 그래프 방법에 비해 노드 간의 전역적, 지역적 관계를 동시에 잘 고려함으로써 검출 성능을 높이는 역할을 한다. 실험에서는 두 종류의 씨드가 오검출 (false positive error) 에러를 줄이는 동시에, 다층 그래프를 통하여 검출 성능이 보

다 개선되는 것을 확인한다. 또한 제안한 방법이 다른 최신 논문의 방법들 보다 다양한 객관적인 평가에서 좋은 성능을 보이는 것을 확인할 수 있다.

본 논문에서는 복수의 이미지에서 공통적으로 중요한 객체를 검출하는 방법 또한 제안한다. 복수 이미지에서의 공통 객체 검출의 경우에는 단일 이미지에서의 중요 객체 검출 방법과는 다르게, 객체의 중요성을 나타내는 정보 이외에도 여러 이미지에서의 공통성을 나타내는 추가적인 정보가 필요하다. 이를 위해 “중요 객체에 대한 유사성” (similarity of saliency, SoS)을 설계한다. 이외의 씨드 추출하는 방법과 그래프에서의 전파 방법 등 기본적인 방법은 단일 이미지에서의 중요 객체 검출 방법과 유사하게 이루어진다. 하지만 여러 이미지에서의 정보를 함께 고려하여 한 그래프 상에서 전파하기 위해서 클러스터 노드(cluster nodes)를 통하여 간접적으로 연결한다. 이는 이미지 내의 객체의 위치 및 크기 차이에 의해 다른 이미지의 노드 간의 직접적인 연관성을 고려하기 힘들기 때문이다. 실험 결과는 공통 중요 객체에 대한 정보가 효과적으로 여러 이미지 노드에 전파되는 것을 확인할 수 있고, 제안하는 방법이 다른 논문들의 방법들과 비교했을 때 객관적, 주관적으로 좋은 성능을 보인다.

끝으로 제안하는 접근법은 목적이 명확한 주제인 피부 검출에도 적용한다. 배경 씨드를 추출하는 과정은 동일하게 하고, 중요 객체를 나타내는 씨드를 추출하는 대신에 씨드를 추출한다. 추출된 두 종류의 씨드를 전파하기 위하여 그래프를 생성하는데, 피부에 관한 정보를 그래프 생성 과정에 이용함으로써 노드들을 보다 효과적으로 연결한다. 제안하는 피부 검출 방법은 다른 논문들의 비해 객관적인 지표에서 높은 검출 성능을 보였다. 이는 제안하는 접근 방법이 피부 검출과 같은 특정한 목적의 검출 방법에도 활용될 수 있음을 시사하고, 다른 주제에도 활용될 수 있을 것이라 기대한다.

주요어: 중요 객체 검출, 공통 주요 객체 검출, 피부 검출, 다중 씨드, 씨드 전파, 다층 그래프

학번: 2012-30239